PCT

wo

!TELLECTUAL PROPERTY ORGANIZATION International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6: C12N 15/31, 15/82, 15/10, 1/21, 5/10,

A01H 5/00, C07K 14/24, A01N 63/02

A2

(11) International Publication Number:

WO 99/42589

US

US

(43) International Publication Date:

26 August 1999 (26.08.99)

(21) International Application Number:

PCT/EP99/01015

(22) International Filing Date:

18 February 1999 (18.02.99)

(30) Priority Data:

09/027,080 60/116,439 20 February 1998 (20.02.98)

20 January 1999 (20.01.99)

(71) Applicant (for all designated States except AT US): NOVAR-TIS AG [CH/CH]; Schwarzwaldallee 215, D-4058 Basel (CH).

(71) Applicant (for AT only): NOVARTIS-ERFINDUNGEN VER-WALTUNGSGESELLSCHAFT MBH [AT/AT]; Brunner Strasse 59, A-1235 Vienna (AT).

(75) Inventors/Applicants (for US only): KRAMER, Vance, Cary [US/US]; 608 Dana Court, Hillsborough, NC 27278 (US). MORGAN, Michael, Kent [US/US]; 5805 Garrett Road, Durham, NC 27707 (US). ANDERSON, Arne, Robert [US/US]: 1005 Green-Pace Road, Zebulon, NC 27597 (US). HART, Hope, Prim [US/US]; 4106 Planters Glen Court, Fuquay-Varina, NC 26526 (US). Warren, Gregory, Wayne [US/US]; 324 Bond Lake Drive, Cary, NC 27513 (US). DUNN, Martha, M. [US/US]; 6201 Oakview Court, Hillsborough, NC 27278 (US). CHEN, Jeng, Shong [-/US]; 302 Orchard Lane, Chapel Hill, NC 27514 (US).

(74) Agent: BECKER, Konrad; Novartis AG, Corporate Intellectual Property, Patent & Trademark Dept., CH-4002 Basel (CH).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: INSECTICIDAL TOXINS FROM PHOTORHABDUS

(57) Abstract

Novel nucleic acid sequences isolated from Photorhabdus luminescens, whose expression results in novel insecticidal toxins, are disclosed herein. The invention also discloses compositions and formulations containing the insecticidal toxins that are capable of controlling insect pests. The invention is further drawn to methods of making the toxins and to methods of using the nucleotide sequences, for example in microorganisms to control insect pests or in transgenic plants to confer insect resistance.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AΤ	Austria	FR	France	LU	Luxembourg	SN	Senegal
ΑU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
вв	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Кепуа	NL	Netherlands	YU	Yugoslavia
СН	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

-1-

INSECTICIDAL TOXINS FROM PHOTORHABDUS

The invention relates to novel toxins from *Photorhabdus luminescens*, nucleic acid sequences whose expression results in said toxins, and methods of making and methods of using the toxins and corresponding nucleic acid sequences to control insects.

Insect pests are a major cause of crop losses. Solely in the US, about \$7.7 billion are lost every year due to infestation by various genera of insects. In addition to losses in field crops, insect pests are also a burden to vegetable and fruit growers, to producers of ornamental flowers, and they are a nuisance to gardeners and home owners.

Insect pests are mainly controlled by intensive applications of chemical insecticides, which are active through inhibition of insect growth, prevention of insect feeding or reproduction, or death of the insects. Good insect control can thus be reached, but these chemicals can sometimes also affect other, beneficial insects. Another problem resulting from the wide use of chemical pesticides is the appearance of resistant insect varieties. This has been partially alleviated by various resistance management strategies, but there is an increasing need for alternative pest control agents. Biological insect control agents, such as Bacillus thuringiensis strains expressing insecticidal toxins like d-endotoxins, have also been applied with satisfactory results, offering an alternative or a complement to chemical insecticides. Recently, the genes coding for some of these d-endotoxins have been isolated and their expression in heterologous hosts have been shown to provide another tool for the control of economically important insect pests. In particular, the expression of insecticidal toxins in transgenic plants, such as Bacillus thuringiensis dendotoxins, has provided efficient protection against selected insect pests, and transgenic plants expressing such toxins have been commercialized, allowing farmers to reduce applications of chemical insect control agents. Yet, even in this case, the development of resistance remains a possibility and only a few specific insect pests are controllable. Consequently, there remains a long-felt but unfulfilled need to discover new and effective insect control agents that provide an economic benefit to farmers and that are environmentally acceptable.

The present invention addresses the need for novel insect control agents. Particularly needed are control agents that are targeted to economically important insect pests and that efficiently control insect strains resistant to existing insect control agents.

-2-

Furthermore, agents whose application minimizes the burden on the environment are desirable.

In the search of novel insect control agents, certain classes of nematodes from the genera *Heterorhabdus* and *Steinernema* are of particular interest because of their insecticidal properties. They kill insect larvae and their offspring feed in the dead larvae. Indeed, the insecticidal activity is due to symbiotic bacteria living in the nematodes. These symbiotic bacteria are *Photorhabdus* in the case of *Heterorhabdus* and *Xenorhabdus* in the case of *Steinernema*.

The present invention is drawn to nucleic acid sequences isolated from *Photorhabdus luminescens*, and sequences substantially similar thereto, whose expression results in toxins that are highly toxic to economically important insect pests, particularly insect pests that infest plants. The invention is further drawn to the toxins resulting from the expression of the nucleic acid sequences, and to compositions and formulations containing the toxins, which are capable of inhibiting the ability of insect pests to survive, grow or reproduce, or of limiting insect-related damage or loss in crop plants. The invention is further drawn to a method of making the toxins and to methods of using the nucleic acid sequences, for example in microorganisms to control insects or in transgenic plants to confer insect resistance, and to a method of using the toxins, and compositions and formulations comprising the toxins, for example applying the toxins or compositions or formulations to insect-infested areas, or to prophylactically treat insect-susceptible areas or plants to confer protection or resistance to the insects.

The novel toxins are highly active against insects. For example, a number of economically important insect pests, such as the Lepidopterans *Plutella xylostella* (Diamondback Moth), *Trichoplusia ni* (Cabbage Looper), *Ostrinia nubilalis* (European Corn Borer), *Heliothis virescens* (Tobacco Budworm), *Helicoverpa zea* (Corn Earworm), *Manduca sexta* (Tobacco Hornworm), *Spodoptera exigua* (Beet Armyworm), and *Spodoptera frugiperda* (Fall Armyworm), as well as the Coleopterans *Diabrotica virgifera virgifera* (Western Corn Rootworm), *Diabrotica undecimpunctata howardi* (Southern Corn Rootworm), and *Leptinotarsa decimlineata* (Colorado Potato Beetle) can be controlled by one or more of the toxins. The toxins can be used in multiple insect control strategies, resulting in maximal efficiency with minimal impact on the environment.

According to one aspect, the present invention provides an isolated nucleic acid molecule comprising: (a) a nucleotide sequence substantially similar to a nucleotide

sequence selected from the group consisting of: nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, nucleotides 4515-9269 of SEQ ID NO:1, nucleotides 15,171-18,035 of SEQ ID NO:11, and nucleotides 31,393-35,838 of SEQ ID NO:11; (b) a nucleotide sequence comprising nucleotides 23,768-31,336 of SEQ ID NO:11; or (c) a nucleotide sequence isocoding with the nucleotide sequence of (a) or (b); wherein expression of the nucleic acid molecule results in at least one toxin that is active against insects.

In one embodiment of this aspect, the nucleotide sequence is isocoding with a nucleotide sequence substantially similar to nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, or nucleotides 4515-9269 of SEQ ID NO:1. Preferably, the nucleotide sequence is substantially similar to nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, or nucleotides 4515-9269 of SEQ ID NO:1. More preferably, the nucleotide sequence encodes an amino acid sequence selected from the group consisting of SEQ ID NO:2-6. Most preferably, the nucleotide sequence comprises nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, or nucleotides 4515-9269 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, or nucleotides 4515-9269 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, or nucleotides 4515-9269 of SEQ ID NO:1.

In another embodiment of this aspect, the nucleotide sequence is isocoding with a nucleotide sequence substantially similar to nucleotides 15,171-18,035 of SEQ ID NO:11. Preferably, the nucleotide sequence is substantially similar to nucleotides 15,171-18,035 of SEQ ID NO:11. More preferably, the nucleotide sequence encodes the amino acid sequence set forth in SEQ ID NO:12. Most preferably, the nucleotide sequence comprises nucleotides 15,171-18,035 of SEQ ID NO:11.

In still another embodiment of this aspect, the nucleotide sequence is isocoding with a nucleotide sequence substantially similar to nucleotides 31,393-35,838 of SEQ ID NO:11. Preferably, the nucleotide sequence is substantially similar to nucleotides 31,393-35,838 of SEQ ID NO:11. More preferably, the nucleotide sequence encodes the amino acid sequence set forth in SEQ ID NO:14. Most preferably, the nucleotide sequence comprises nucleotides 31,393-35,838 of SEQ ID NO:11.

- 4 -

In yet another embodiment of this aspect, the nucleotide sequence encodes the amino acid sequence set forth in SEQ ID NO:13, and preferably comprises nucleotides 23,768-31,336 of SEQ ID NO:11.

In one embodiment, the nucleotide sequence of the invention comprises the approximately 9.7 kb DNA fragment harbored in *E. coli* strain DH5a, designated as NRRL accession number B-21835.

In another embodiment, the nucleotide sequence of the invention comprises the approximately 38 kb DNA fragment harbored in *E. coli* strain DH5a, designated as NRRL accession number B-30077.

In still another embodiment, the nucleotide sequence of the invention comprises the approximately 22.2 kb DNA fragment harbored in *E. coli* strain DH5a, designated as NRRL accession number B-30078.

According to one embodiment of the invention, the toxins resulting from expression of the nucleic acid molecules of the invention have activity against Lepidopteran insects. Preferably, according to this embodiment, the toxins have activity against *Plutella xylostella* (Diamondback Moth), *Trichoplusia ni* (Cabbage Looper), *Ostrinia nubilalis* (European Corn Borer), *Heliothis virescens* (Tobacco Budworm), *Helicoverpa zea* (Corn Earworm), *Spodoptera exigua* (Beet Armyworm), and *Spodoptera frugiperda* (Fall Armyworm).

According to another embodiment of the invention, the toxins resulting from expression of the nucleic acid molecule of the invention have activity against Lepidopteran and Coleopteran insects. Preferably, according to this embodiment, the toxins have insecticidal activity against *Plutella xylostella* (Diamondback Moth), *Ostrinia nubilalis* (European Corn Borer), and *Manduca sexta* (Tobacco Hornworm), *Diabrotica virgifera virgifera* (Western Corn Rootworm), *Diabrotica undecimpunctata howardi* (Southern Corn Rootworm), and *Leptinotarsa decimlineata* (Colorado Potato Beetle).

In another aspect, the present invention provides an isolated nucleic acid molecule comprising a 20 base pair nucleotide portion identical in sequence to a consecutive 20 base pair nucleotide portion of a nucleotide sequence selected from the group consisting of: nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, nucleotides 4515-9269 of SEQ ID NO:1, nucleotides 15,171-18,035 of SEQ ID NO:11, and nucleotides 31,393-35,838 of SEQ ID NO:11, wherein expression of the nucleic acid molecule results in at least one toxin that is active against insects.

In one embodiment of this aspect, the isolated nucleic acid molecule of the invention comprises a 20 base pair nucleotide portion identical in sequence to a consecutive 20 base pair nucleotide portion of nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, or nucleotides 4515-9269 of SEQ ID NO:1.

In another embodiment of this aspect, the isolated nucleic acid molecule of the invention comprises a 20 base pair nucleotide portion identical in sequence to a consecutive 20 base pair nucleotide portion of nucleotides 15,171-18,035 of SEQ ID NO:11.

In still another embodiment of this aspect, the isolated nucleic acid molecule of the invention comprises a 20 base pair nucleotide portion identical in sequence to a consecutive 20 base pair nucleotide portion of nucleotides 31,393-35,838 of SEQ ID NO:11.

In a further aspect, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence from *Photorhabdus luminescens* selected from the group consisting of: nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, nucleotides 4515-9269 of SEQ ID NO:1, nucleotides 66-1898 of SEQ ID NO:11, nucleotides 2416-9909 of SEQ ID NO:11, the complement of nucleotides 2817-3395 of SEQ ID NO:11, nucleotides 9966-14,633 of SEQ ID NO:11, nucleotides 14,699-15,007 of SEQ ID NO:11, nucleotides 15,171-18,035 of SEQ ID NO:11, the complement of nucleotides 17,072-17,398 of SEQ ID NO:11, the complement of nucleotides 18,235-19,167 of SEQ ID NO:11, the complement of nucleotides 20,217-20,963 of SEQ ID NO:11, the complement of nucleotides 22,172-23,086 of SEQ ID NO:11, nucleotides 23,768-31,336 of SEQ ID NO:11, nucleotides 31,393-35,838 of SEQ ID NO:11, the complement of nucleotides 35,383-35,709 of SEQ ID NO:11, the complement of nucleotides 36,032-36,661 of SEQ ID NO:11, and the complement of nucleotides 36,654-37,781 of SEQ ID NO:11.

The present invention also provides a chimeric gene comprising a heterologous promoter sequence operatively linked to the nucleic acid molecule of the invention. Further, the present invention provides a recombinant vector comprising such a chimeric gene. Still further, the present invention provides a host cell comprising such a chimeric gene. A host cell according to this aspect of the invention may be a bacterial cell, a yeast cell, or a plant

cell, preferably a plant cell. Even further, the present invention provides a plant comprising such a plant cell. Preferably, the plant is maize.

In yet another aspect, the present invention provides toxins produced by the expression of DNA molecules of the present invention.

According to one embodiment, the toxins of the invention have activity against Lepidopteran insects, preferably against *Plutella xylostella* (Diamondback Moth), *Trichoplusia ni* (Cabbage Looper), *Ostrinia nubilalis* (European Corn Borer), *Heliothis virescens* (Tobacco Budworm), *Helicoverpa zea* (Corn Earworm), *Spodoptera exigua* (Beet Armyworm), and *Spodoptera frugiperda* (Fall Armyworm).

According to another embodiment, the toxins of the invention have activity against Lepidopteran and Coleopteran insects, preferably against *Plutella xylostella* (Diamondback Moth), *Ostrinia nubilalis* (European Corn Borer), and *Manduca sexta* (Tobacco Hornworm), *Diabrotica virgifera virgifera* (Western Corn Rootworm), *Diabrotica undecimpunctata howardi* (Southern Corn Rootworm), and *Leptinotarsa decimlineata* (Colorado Potato Beetle).

In one embodiment, the toxins are produced by the *E. coli* strain designated as NRRL accession number B-21835.

In another embodiment, the toxins are produced by *E. coli* strain designated as NRRL accession number B-30077.

In still another embodiment, the toxins are produced by *E. coli* strain designated as NRRL accession number B-30078.

In one embodiment, a toxin of the invention comprises an amino acid sequence selected from the group consisting of: SEQ ID NOs:2-6.

In another embodiment, a toxin of the invention comprises an amino acid sequence selected from the group consisting of: SEQ ID NOs:12-14.

The present invention also provides a composition comprising an insecticidally effective amount of a toxin according to the invention.

In another aspect, the present invention provides a method of producing a toxin that is active against insects, comprising: (a) obtaining a host cell comprising a chimeric gene, which itself comprises a heterologous promoter sequence operatively linked to the nucleic acid molecule of the invention; and (b) expressing the nucleic acid molecule in the cell, which results in at least one toxin that is active against insects.

In a further aspect, the present invention provides a method of producing an insect-resistant plant, comprising introducing a nucleic acid molecule of the invention into the plant, wherein the nucleic acid molecule is expressible in the plant in an effective amount to control insects. According to one embodiment, the insects are Lepidopteran insects, preferably selected from the group consisting of: *Plutella xylostella* (Diamondback Moth), *Trichoplusia ni* (Cabbage Looper), *Ostrinia nubilalis* (European Corn Borer), *Heliothis virescens* (Tobacco Budworm), *Helicoverpa zea* (Corn Earworm), *Spodoptera exigua* (Beet Armyworm), and *Spodoptera frugiperda* (Fall Armyworm). According to another embodiment, the insects are Lepidopteran and Coleopteran insects, preferably selected from the group consisting of: *Plutella xylostella* (Diamondback Moth), *Ostrinia nubilalis* (European Corn Borer), and *Manduca sexta* (Tobacco Hornworm), *Diabrotica virgifera virgifera* (Western Corn Rootworm), *Diabrotica undecimpunctata howardi* (Southern Corn Rootworm), and *Leptinotarsa. decimlineata* (Colorado Potato Beetle).

In still a further aspect, the present invention provides a method of controlling insects comprising delivering to the insects an effective amount of a toxin according to the present invention. According to one embodiment, the insects are Lepidopteran insects, preferably selected from the group consisting of: *Plutella xylostella* (Diamondback Moth), *Trichoplusia ni* (Cabbage Looper), *Ostrinia nubilalis* (European Corn Borer), *Heliothis virescens* (Tobacco Budworm), *Helicoverpa zea* (Corn Earworm), *Spodoptera exigua* (Beet Armyworm), and *Spodoptera frugiperda* (Fall Armyworm). According to another embodiment, the insects are Lepidopteran and Coleopteran insects, preferably selected from the group consisting of: *Plutella xylostella* (Diamondback Moth), *Ostrinia nubilalis* (European Corn Borer), and *Manduca sexta* (Tobacco Hornworm), *Diabrotica virgifera virgifera* (Western Corn Rootworm), *Diabrotica undecimpunctata howardi* (Southern Corn Rootworm), and *Leptinotarsa decimlineata* (Colorado Potato Beetle). Preferably, the toxin is delivered to the insects orally.

Yet another aspect of the present invention is the provision of a method for mutagenizing a nucleic acid molecule according to the present invention, wherein the nucleic acid molecule has been cleaved into population of double-stranded random fragments of a desired size, comprising: (a) adding to the population of double-stranded random fragments one or more single- or double-stranded oligonucleotides, wherein the oligonucleotides each comprise an area of identity and an area of heterology to a double-stranded template polynucleotide; (b) denaturing the resultant mixture of double-stranded

-8-

random fragments and oligonucleotides into single-stranded fragments; (c) incubating the resultant population of single-stranded fragments with a polymerase under conditions which result in the annealing of the single-stranded fragments at the areas of identity to form pairs of annealed fragments, the areas of identity being sufficient for one member of a pair to prime replication of the other, thereby forming a mutagenized double-stranded polynucleotide; and (d) repeating the second and third steps for at least two further cycles, wherein the resultant mixture in the second step of a further cycle includes the mutagenized double-stranded polynucleotide from the third step of the previous cycle, and wherein the further cycle forms a further mutagenized double-stranded polynucleotide.

Other aspects and advantages of the present invention will become apparent to those skilled in the art from a study of the following description of the invention and non-limiting examples.

DEFINITIONS

"Activity" of the toxins of the invention is meant that the toxins function as orally active insect control agents, have a toxic effect, or are able to disrupt or deter insect feeding, which may or may not cause death of the insect. When a toxin of the invention is delivered to the insect, the result is typically death of the insect, or the insect does not feed upon the source that makes the toxin available to the insect.

"Associated with / operatively linked" refer to two nucleic acid sequences that are related physically or functionally. For example, a promoter or regulatory DNA sequence is said to be "associated with" a DNA sequence that codes for an RNA or a protein if the two sequences are operatively linked, or situated such that the regulator DNA sequence will affect the expression level of the coding or structural DNA sequence.

A "chimeric gene" is a recombinant nucleic acid sequence in which a promoter or regulatory nucleic acid sequence is operatively linked to, or associated with, a nucleic acid sequence that codes for an mRNA or which is expressed as a protein, such that the regulator nucleic acid sequence is able to regulate transcription or expression of the associated nucleic acid sequence. The regulator nucleic acid sequence of the chimeric gene is not normally operatively linked to the associated nucleic acid sequence as found in nature.

A "coding sequence" is a nucleic acid sequence that is transcribed into RNA such as mRNA, rRNA, tRNA, snRNA, sense RNA or antisense RNA. Preferably the RNA is then translated in an organism to produce a protein.

To "control" insects means to inhibit, through a toxic effect, the ability of insect pests to survive, grow, feed, and/or reproduce, or to limit insect-related damage or loss in crop plants. To "control" insects may or may not mean killing the insects, although it preferably means killing the insects.

To "deliver" a toxin means that the toxin comes in contact with an insect, resulting in toxic effect and control of the insect. The toxin can be delivered in many recognized ways, e.g., orally by ingestion by the insect or by contact with the insect via transgenic plant expression, formulated protein composition(s), sprayable protein composition(s), a bait matrix, or any other art-recognized toxin delivery system.

"Expression cassette" as used herein means a nucleic acid sequence capable of directing expression of a particular nucleotide sequence in an appropriate host cell, comprising a promoter operably linked to the nucleotide sequence of interest which is operably linked to termination signals. It also typically comprises sequences required for proper translation of the nucleotide sequence. The expression cassette comprising the nucleotide sequence of interest may be chimeric, meaning that at least one of its components is heterologous with respect to at least one of its other components. The expression cassette may also be one which is naturally occurring but has been obtained in a recombinant form useful for heterologous expression. Typically, however, the expression cassette is heterologous with respect to the host, i.e., the particular nucleic acid sequence of the expression cassette does not occur naturally in the host cell and must have been introduced into the host cell or an ancestor of the host cell by a transformation event. The expression of the nucleotide sequence in the expression cassette may be under the control of a constitutive promoter or of an inducible promoter which initiates transcription only when the host cell is exposed to some particular external stimulus. In the case of a multicellular organism, such as a plant, the promoter can also be specific to a particular tissue, or organ, or stage of development.

A "gene" is a defined region that is located within a genome and that, besides the aforementioned coding nucleic acid sequence, comprises other, primarily regulatory, nucleic acid sequences responsible for the control of the expression, that is to say the transcription and translation, of the coding portion. A gene may also comprise other 5' and 3'

untranslated sequences and termination sequences. Further elements that may be present are, for example, introns.

"Gene of interest" refers to any gene which, when transferred to a plant, confers upon the plant a desired characteristic such as antibiotic resistance, virus resistance, insect resistance, disease resistance, or resistance to other pests, herbicide tolerance, improved nutritional value, improved performance in an industrial process or altered reproductive capability. The "gene of interest" may also be one that is transferred to plants for the production of commercially valuable enzymes or metabolites in the plant.

A "heterologous" nucleic acid sequence is a nucleic acid sequence not naturally associated with a host cell into which it is introduced, including non-naturally occurring multiple copies of a naturally occurring nucleic acid sequence.

A "homologous" nucleic acid sequence is a nucleic acid sequence naturally associated with a host cell into which it is introduced.

"Homologous recombination" is the reciprocal exchange of nucleic acid fragments between homologous nucleic acid molecules.

"Insecticidal" is defined as a toxic biological activity capable of controlling insects, preferably by killing them.

A nucleic acid sequence is "isocoding with" a reference nucleic acid sequence when the nucleic acid sequence encodes a polypeptide having the same amino acid sequence as the polypeptide encoded by the reference nucleic acid sequence.

An "isolated" nucleic acid molecule or an isolated enzyme is a nucleic acid molecule or enzyme that, by the hand of man, exists apart from its native environment and is therefore not a product of nature. An isolated nucleic acid molecule or enzyme may exist in a purified form or may exist in a non-native environment such as, for example, a recombinant host cell.

A "nucleic acid molecule" or "nucleic acid sequence" is a linear segment of single- or double-stranded DNA or RNA that can be isolated from any source. In the context of the present invention, the nucleic acid molecule is preferably a segment of DNA.

"ORF" means open reading frame.

A "plant" is any plant at any stage of development, particularly a seed plant.

A "plant cell" is a structural and physiological unit of a plant, comprising a protoplast and a cell wall. The plant cell may be in form of an isolated single cell or a cultured cell, or as a part of higher organized unit such as, for example, plant tissue, a plant organ, or a whole plant.

WO 99/42589

PCT/EP99/01015

- 11 -

"Plant cell culture" means cultures of plant units such as, for example, protoplasts, cell culture cells, cells in plant tissues, pollen, pollen tubes, ovules, embryo sacs, zygotes and embryos at various stages of development.

"Plant material" refers to leaves, stems, roots, flowers or flower parts, fruits, pollen, egg cells, zygotes, seeds, cuttings, cell or tissue cultures, or any other part or product of a plant.

A "plant organ" is a distinct and visibly structured and differentiated part of a plant such as a root, stem, leaf, flower bud, or embryo.

"Plant tissue" as used herein means a group of plant cells organized into a structural and functional unit. Any tissue of a plant *in planta* or in culture is included. This term includes, but is not limited to, whole plants, plant organs, plant seeds, tissue culture and any groups of plant cells organized into structural and/or functional units. The use of this term in conjunction with, or in the absence of, any specific type of plant tissue as listed above or otherwise embraced by this definition is not intended to be exclusive of any other type of plant tissue.

A "promoter" is an untranslated DNA sequence upstream of the coding region that contains the binding site for RNA polymerase II and initiates transcription of the DNA. The promoter region may also include other elements that act as regulators of gene expression.

A "protoplast" is an isolated plant cell without a cell wall or with only parts of the cell wall.

"Regulatory elements" refer to sequences involved in controlling the expression of a nucleotide sequence. Regulatory elements comprise a promoter operably linked to the nucleotide sequence of interest and termination signals. They also typically encompass sequences required for proper translation of the nucleotide sequence.

In its broadest sense, the term "substantially similar", when used herein with respect to a nucleotide sequence, means a nucleotide sequence corresponding to a reference nucleotide sequence, wherein the corresponding sequence encodes a polypeptide having substantially the same structure and function as the polypeptide encoded by the reference nucleotide sequence, e.g. where only changes in amino acids not affecting the polypeptide function occur. Desirably the substantially similar nucleotide sequence encodes the polypeptide encoded by the reference nucleotide sequence. The percentage of identity between the substantially similar nucleotide sequence and the reference nucleotide sequence desirably is at least 80%, more desirably at least 85%, preferably at least 90%, more preferably at least 95%, still more preferably at least 99%. A nucleotide sequence

"substantially similar" to reference nucleotide sequence hybridizes to the reference nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C, more desirably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1% SDS at 50°C, more desirably still in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C, preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C, more preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C.

"Synthetic" refers to a nucleotide sequence comprising structural characters that are not present in the natural sequence. For example, an artificial sequence that resembles more closely the G+C content and the normal codon distribution of dicot and/or monocot genes is said to be synthetic.

"Transformation" is a process for introducing heterologous nucleic acid into a host cell or organism. In particular, "transformation" means the stable integration of a DNA molecule into the genome of an organism of interest.

"Transformed / transgenic / recombinant" refer to a host organism such as a bacterium or a plant into which a heterologous nucleic acid molecule has been introduced. The nucleic acid molecule can be stably integrated into the genome of the host or the nucleic acid molecule can also be present as an extrachromosomal molecule. Such an extrachromosomal molecule can be auto-replicating. Transformed cells, tissues, or plants are understood to encompass not only the end product of a transformation process, but also transgenic progeny thereof. A "non-transformed", "non-transgenic", or "non-recombinant" host refers to a wild-type organism, e.g., a bacterium or plant, which does not contain the heterologous nucleic acid molecule.

Nucleotides are indicated by their bases by the following standard abbreviations: adenine (A), cytosine (C), thymine (T), and guanine (G). Amino acids are likewise indicated by the following standard abbreviations: alanine (Ala; A), arginine (Arg; R), asparagine (Asn; N), aspartic acid (Asp; D), cysteine (Cys; C), glutamine (Gln; Q), glutamic acid (Glu; E), glycine (Gly; G), histidine (His; H), isoleucine (Ile; I), leucine (Leu; L), lysine (Lys; K), methionine (Met; M), phenylalanine (Phe; F), proline (Pro; P), serine (Ser; S), threonine (Thr; T), tryptophan (Trp; W), tyrosine (Tyr; Y), and valine (Val; V). Furthermore, (Xaa; X) represents any amino acid.

BRIEF DESCRIPTION OF THE SEQUENCES IN THE SEQUENCE LISTING

SEQ ID NO:1 is the sequence of the approximately 9.7 kb DNA fragment comprised in pCIB9359-7 which comprises the following ORFs at the specified nucleotide positions:

<u>Name</u>	Start	<u>End</u>
orf1	412	1665
orf2	1686	2447
orf3	2758	3318
orf4	3342	4118
orf5	4515	9269

SEQ ID NO:2 is the sequence of the ~46.4 kDa protein encoded by orf1 of SEQ ID NO:1.

SEQ ID NO:3 is the sequence of the ~28.1 kDa protein encoded by orf2 of SEQ ID NO:1.

SEQ ID NO:4 is the sequence of the ~20.7 kDa protein encoded by orf3 of SEQ ID NO:1.

SEQ ID NO:5 is the sequence of the ~28.7 kDa protein encoded by orf4 of SEQ ID NO:1.

SEQ ID NO:6 is the sequence of the ~176 kDa protein encoded by orf5 of SEQ ID NO:1.

SEQ ID NOs:7-10 are oligonucleotides.

SEQ ID NO:11 is the sequence of the approximately 38 kb DNA fragment comprised in pNOV2400, which comprises the following ORFs at the specified nucleotide positions (descending numbers and "C" indicates that the ORF is on the complementary strand):

<u>Name</u>	Start	<u>End</u>	
orf7	66	1898	(partial sequence)
hph3	2416	9909	
orf18	3395	2817	С
orf4	9966	14,633	
orf19	14,699	15,007	
orf5	15,171	18,035	
orf22	17,398	17,072	С
orf10	19,167	18,235	С
orf14	20,116	19,385	С
orf13	20,963	20,217	С
orf11	23,086	22,172	С
hph2	23,768	31,336	
orf2	31,393	35,838	

- 14 -

orf21 35,709 35,383 C orf16 36,661 36,032 C orf8 37,781 36,654 C

SEQ ID NO:11 also includes the following restriction sites, some of which are used in the subcloning steps set forth in Example 17:

Restriction Site	Nucleotide Position(s)		
Acdll	2835		
<i>Bam</i> HI	18,915		
<i>Bsm</i> Bl	11,350		
<i>Bst</i> 11071	29,684		
<i>Eag</i> l	13,590; 31,481		
Eco721	34,474		
Mlul	2444; 5116; 9327; 26,204		
Notl	13,589		
<i>Pac</i> l	9915; 23,353; 37,888		
Pvul	8816		
Sapl	35,248		
SexAI	28,946		
Sgfl	8815		
Spel	2157; 3769; 7831; 11,168		
Sphi	755		
Stul	35,690		
Tth1111	21,443		

SEQ ID NO:12 is the sequence of the protein encoded by orf5 of SEQ ID NO:11.

SEQ ID NO:13 is the sequence of the protein encoded by hph2 of SEQ ID NO:11.

SEQ ID NO:14 is the sequence of the protein encoded by orf2 of SEQ ID NO:11.

SEQ ID NOs:15-22 are oligonucleotides.

DEPOSITS

The following material has been deposited with the Agricultural Research Service, Patent Culture Collection (NRRL), 1815 North University Street, Peoria, Illinois 61604, under the terms of the Budapest Treaty on the International Recognition of the Deposit of

Microorganisms for the Purposes of Patent Procedure. All restrictions on the availability of

the deposited material will be irrevocably removed upon the granting of a patent.

<u>Clone</u>	Accession Number	Date of Deposit
pCIB9359-7	NRRL B-21835	September 17, 1997
pNOV2400	NRRL B-30077	December 3, 1998
pNOV1001	NRRL B-30078	December 3, 1998

Novel Nucleic Acid Sequences whose Expression Results in Insecticidal Toxins

This invention relates to nucleic acid sequences whose expression results in novel toxins, and to the making and using of the toxins to control insect pests. The nucleic acid sequences are derived from Photorhabdus luminescens, a member of Enteropacteriaceae family. P. luminescens is a symbiotic bacterium of nematodes of the genus Heterorhabditis. The nematodes colonize insect larva, kill them, and their offspring feed on the dead larvae. The insecticidal activity is actually produced by the symbiotic P. luminescens bacteria. The inventors are the first to isolate the nucleic acid sequences of the present invention from P. luminescens (ATCC strain number 29999). The expression of the nucleic acid sequences of the present invention results in toxins that can be used to control Lepidopteran insects such as Plutella xylostella (Diamondback Moth), Trichoplusia ni (Cabbage Looper), Ostrinia nubilalis (European Corn Borer), Heliothis virescens (Tobacco Budworm), Helicoverpa zea (Corn Earworm), Manduca sexta (Tobacco Hornworm), Spodoptera exigua (Beet Armyworm), and Spodoptera frugiperda (Fall Armyworm), as well as Coleopteran insects such as Diabrotica virgifera virgifera (Western Corn Rootworm), Diabrotica undecimpunctata howardi (Southern Corn Rootworm), Diabrotica longicornis barberi (Northern Corn Rootworm), and Leptinotarsa decimlineata (Colorado Potato Beetle).

In one preferred embodiment, the invention encompasses an isolated nucleic acid molecule comprising a nucleotide sequence substantially similar to the approximately 9.7 kb nucleic acid sequence set forth in SEQ ID NO:1, whose expression results in insect control activity (further illustrated in Examples 1-11). Five open reading frames (ORFs) are present in the nucleic acid sequence set forth in SEQ ID NO:1, coding for proteins of predicted sizes 45 kDa, 28 kDa, 21 kDA, 29 kDa, and 176 kDa. The five ORFs are arranged in an operon-like structure. When expressed in a heterologous host, the ~ 9.7 kb DNA fragment from *P*.

WO 99/42589 PCT/EP99/01015 - 16 -

luminescens results in insect control activity against Lepidopterans such as *Plutella xylostella* (Diamondback Moth), *Trichoplusia ni* (Cabbage Looper), *Ostrinia nubilalis* (European Corn Borer), *Heliothis virescens* (Tobacco Budworm), *Helicoverpa zea* (Corn Earworm), *Spodoptera exigua* (Beet Armyworm), and *Spodoptera frugiperda* (Fall Armyworm), showing that expression of the ~ 9.7 kb nucleotide sequence set forth in SEQ ID NO:1 is necessary and sufficient for such insect control activity. In a preferred embodiment, the invention encompasses a DNA molecule, whose expression results in an insecticidal toxin, which is deposited in the *E. coli* strain pCIB9359-7 (NRRL accession number B-21835).

In another preferred embodiment, the invention encompasses an isolated nucleic acid molecule comprising a nucleotide sequence substantially similar to the approximately 38 kb nucleic acid fragment set forth in SEQ ID NO:11 and deposited in the E. coli strain pNOV2400 (NRRL accession number B-30077), whose expression results in insect control activity (see Examples 12-18). In a more preferred embodiment, the invention encompasses an isolated nucleic acid molecule comprising a nucleotide sequence substantially similar to the ~ 22 kb DNA fragment deposited in the E. coli strain pNOV1001 (NRRL accession number B-30078), whose expression results in insect control activity. In a most preferred embodiment, the invention encompasses isolated nucleic acid molecules comprising nucleotide sequences substantially similar to the three ORFs corresponding to nucleotides 23,768-31,336 (hph2), 31,393-35,838 (orf2), and 15,171-18,035 (orf5) of the DNA fragment set forth in SEQ ID NO:11, as well as the proteins encoded thereby. When co-expressed in a heterologous host, these three ORFs result in insect control activity against Lepidopterans such as Plutella xylostella (Diamondback Moth), Ostrinia nubilalis (European Corn Borer), and Manduca sexta (Tobacco Hornworm), as well as against Coleopterans such as Diabrotica virgifera virgifera (Western Corn Rootworm), Diabrotica undecimpunctata howardi (Southern Corn Rootworm), and Leptinotarsa decimlineata (Colorado Potato Beetle), showing that co-expression of these three ORFs (hph2, orf2, and orf5) is necessary and sufficient for such insect control activity.

The present invention also encompasses recombinant vectors comprising the nucleic acid sequences of this invention. In such vectors, the nucleic acid sequences are preferably comprised in expression cassettes comprising regulatory elements for expression of the nucleotide sequences in a host cell capable of expressing the nucleotide sequences. Such regulatory elements usually comprise promoter and termination signals and preferably also

comprise elements allowing efficient translation of polypeptides encoded by the nucleic acid sequences of the present invention. Vectors comprising the nucleic acid sequences are usually capable of replication in particular host cells, preferably as extrachromosomal molecules, and are therefore used to amplify the nucleic acid sequences of this invention in the host cells. In one embodiment, host cells for such vectors are microorganisms, such as bacteria, in particular E.coli. In another embodiment, host cells for such recombinant vectors are endophytes or epiphytes. A preferred host cell for such vectors is a eukaryotic cell, such as a yeast, a plant cell, or an insect cell. Plant cells such as maize cells are most preferred host cells. In another preferred embodiment, such vectors are viral vectors and are used for replication of the nucleotide sequences in particular host cells, e.g. insect cells or plant cells. Recombinant vectors are also used for transformation of the nucleotide sequences of this invention into host cells, whereby the nucleotide sequences are stably integrated into the DNA of such host cells. In one, such host cells are prokaryotic cells. In a preferred embodiment, such host cells are eukaryotic cells, such as yeast cells, insect cells, or plant cells. In a most preferred embodiment, the host cells are plant cells, such as maize cells.

In preferred embodiments, the insecticidal toxins of the invention each comprise at least one polypeptide encoded by a nucleotide sequence of the invention. In another preferred embodiment, the insecticidal toxins are produced from a purified strain of *P. luminescens*, such the strain with ATTC accession number 29999. The toxins of the present invention have insect control activity when tested against insect pests in bioassays; and these properties of the insecticidal toxins are further illustrated in Examples 1-18. The insecticidal toxins desribed in the present invention are further characterized in that their molecular weights are larger than 6,000, as found by size fractionation experiments. The insecticidal toxins retain full insectidical activity after being stored at 4°C for 2 weeks. One is also shown to retain its full insecticidal activity after being freeze-dried and stored at 22°C for 2 weeks. However, the insecticidal toxins of the invention lose their insecticidal activity after incubation for 5 minutes at 100°C.

In further embodiments, the nucleotide sequences of the invention can be modified by incorporation of random mutations in a technique known as *in-vitro* recombination or DNA shuffling. This technique is described in Stemmer et al., Nature 370: 389-391 (1994) and US Patent 5,605,793, which are incorporated herein by reference. Millions of mutant copies of a nucleotide sequence are produced based on an original nucleotide sequence of

- 18 -

this invention and variants with improved properties, such as increased insecticidal activity, enhanced stability, or different specificity or range of target insect pests are recovered. The method encompasses forming a mutagenized double-stranded polynucleotide from a template double-stranded polynucleotide comprising a nucleotide sequence of this invention, wherein the template double-stranded polynucleotide has been cleaved into double-stranded-random fragments of a desired size, and comprises the steps of adding to the resultant population of double-stranded random fragments one or more single or double-stranded oligonucleotides, wherein said oligonucleotides comprise an area of identity and an area of heterology to the double-stranded template polynucleotide; denaturing the resultant mixture of double-stranded random fragments and oligonucleotides into single-stranded fragments; incubating the resultant population of single-stranded fragments with a polymerase under conditions which result in the annealing of said singlestranded fragments at said areas of identity to form pairs of annealed fragments, said areas of identity being sufficient for one member of a pair to prime replication of the other, thereby forming a mutagenized double-stranded polynucleotide; and repeating the second and third steps for at least two further cycles, wherein the resultant mixture in the second step of a further cycle includes the mutagenized double-stranded polynucleotide from the third step of the previous cycle, and the further cycle forms a further mutagenized double-stranded polynucleotide. In a preferred embodiment, the concentration of a single species of doublestranded random fragment in the population of double-stranded random fragments is less than 1% by weight of the total DNA. In a further preferred embodiment, the template double-stranded polynucleotide comprises at least about 100 species of polynucleotides. In another preferred embodiment, the size of the double-stranded random fragments is from about 5 bp to 5 kb. In a further preferred embodiment, the fourth step of the method comprises repeating the second and the third steps for at least 10 cycles.

Expression of the Nucleotide Sequences in Heterologous Microbial Hosts

As biological insect control agents, the insecticidal toxins are produced by expression of the nucleotide sequences in heterologous host cells capable of expressing the nucleotide sequences. In a first embodiment, *P. luminescens* cells comprising modifications of at least one nucleotide sequence of this invention at its chromosomal location are described. Such modifications encompass mutations or deletions of existing regulatory elements, thus leading to altered expression of the nucleotide sequence, or the incorporation of new regulatory elements controlling the expression of the nucleotide sequence. In another

embodiment, additional copies of one or more of the nucleotide sequences are added to *P. luminescens* cells either by insertion into the chromosome or by introduction of extrachromosomally replicating molecules containing the nucleotide sequences.

In another embodiment, at least one of the nucleotide sequences of the invention is inserted into an appropriate expression cassette, comprising a promoter and termination signals. Expression of the nucleotide sequence is constitutive, or an inducible promoter responding to various types of stimuli to initiate transcription is used. In a preferred embodiment, the cell in which the toxin is expressed is a microorganism, such as a virus, a bacteria, or a fungus. In a preferred embodiment, a virus, such as a baculovirus, contains a nucleotide sequence of the invention in its genome and expresses large amounts of the corresponding insecticidal toxin after infection of appropriate eukaryotic cells that are suitable for virus replication and expression of the nucleotide sequence. The insecticidal toxin thus produced is used as an insecticidal agent. Alternatively, baculoviruses engineered to include the nucleotide sequence are used to infect insects *in-vivo* and kill them either by expression of the insecticidal toxin or by a combination of viral infection and expression of the insecticidal toxin.

Bacterial cells are also hosts for the expression of the nucleotide sequences of the invention. In a preferred embodiment, non-pathogenic symbiotic bacteria, which are able to live and replicate within plant tissues, so-called endophytes, or non-pathogenic symbiotic bacteria, which are capable of colonizing the phyllosphere or the rhizosphere, so-called epiphytes, are used. Such bacteria include bacteria of the genera Agrobacterium, Alcaligenes, Azospirillum, Azotobacter, Bacillus, Clavibacter, Enterobacter, Erwinia, Flavobacter, Klebsiella, Pseudomonas, Rhizobium, Serratia, Streptomyces and Xanthomonas. Symbiotic fungi, such as Trichoderma and Gliocladium are also possible hosts for expression of the inventive nucleotide sequences for the same purpose.

Techniques for these genetic manipulations are specific for the different available hosts and are known in the art. For example, the expression vectors pKK223-3 and pKK223-2 can be used to express heterologous genes in *E. coli*, either in transcriptional or translational fusion, behind the *tac or trc* promoter. For the expression of operons encoding multiple ORFs, the simplest procedure is to insert the operon into a vector such as pKK223-3 in transcriptional fusion, allowing the cognate ribosome binding site of the heterologous genes to be used. Techniques for overexpression in gram-positive species such as *Bacillus* are also known in the art and can be used in the context of this invention (Quax *et al. In.:*

- 20 -

Industrial Microorganisms: Basic and Applied Molecular Genetics, *Eds.* Baltz *et al.*, American Society for Microbiology, Washington (1993)). Alternate systems for overexpression rely for example, on yeast vectors and include the use of *Pichia*, *Saccharomyces* and *Kluyveromyces* (Sreekrishna, *In*: Industrial microorganisms: basic and applied molecular genetics, Baltz, Hegeman, and Skatrud *eds.*, American Society for Microbiology, Washington (1993); Dequin & Barre, Biotechnology 12:173-177 (1994); van den Berg *et al.*, Biotechnology 8:135-139 (1990)).

In another preferred embodiment, at least one of the described nucleotide sequences is transferred to and expressed in *Pseudomonas fluorescens* strain CGA267356 (described in the published application EU 0 472 494 and in WO 94/01561) which has biocontrol characteristics. In another preferred embodiment, a nucleotide sequence of the invention is transferred to *Pseudomonas aureofaciens* strain 30-84 which also has biocontrol characteristics. Expression in heterologous biocontrol strains requires the selection of vectors appropriate for replication in the chosen host and a suitable choice of promoter. Techniques are well known in the art for expression in gram-negative and grampositive bacteria and fungi.

Expression of the Nucleotide Sequences in Plant Tissue

In a particularly preferred embodiment, at least one of the insecticidal toxins of the invention is expressed in a higher organism, e.g., a plant. In this case, transgenic plants expressing effective amounts of the toxins protect themselves from insect pests. When the insect starts feeding on such a transgenic plant, it also ingests the expressed toxins. This will deter the insect from further biting into the plant tissue or may even harm or kill the insect. A nucleotide sequence of the present invention is inserted into an expression cassette, which is then preferably stably integrated in the genome of said plant. In another preferred embodiment, the nucleotide sequence is included in a non-pathogenic self-replicating virus. Plants transformed in accordance with the present invention may be monocots or dicots and include, but are not limited to, maize, wheat, barley, rye, sweet potato, bean, pea, chicory, lettuce, cabbage, cauliflower, broccoli, turnip, radish, spinach, asparagus, onion, garlic, pepper, celery, squash, pumpkin, hemp, zucchini, apple, pear, quince, melon, plum, cherry, peach, nectarine, apricot, strawberry, grape, raspberry, blackberry, pineapple, avocado, papaya, mango, banana, soybean, tomato, sorghum, sugarcane, sugarbeet, sunflower, rapeseed, clover, tobacco, carrot, cotton, alfalfa, rice,

- 21 -

potato, eggplant, cucumber, *Arabidopsis*, and woody plants such as coniferous and deciduous trees.

Once a desired nucleotide sequence has been transformed into a particular plant species, it may be propagated in that species or moved into other varieties of the same species, particularly including commercial varieties, using traditional breeding techniques.

A nucleotide sequence of this invention is preferably expressed in transgenic plants, thus causing the biosynthesis of the corresponding toxin in the transgenic plants. In this way, transgenic plants with enhanced resistance to insects are generated. For their expression in transgenic plants, the nucleotide sequences of the invention may require modification and optimization. Although in many cases genes from microbial organisms can be expressed in plants at high levels without modification, low expression in transgenic plants may result from microbial nucleotide sequences having codons that are not preferred in plants. It is known in the art that all organisms have specific preferences for codon usage, and the codons of the nucleotide sequences described in this invention can be changed to conform with plant preferences, while maintaining the amino acids encoded thereby. Furthermore, high expression in plants is best achieved from coding sequences that have at least 35% about GC content, preferably more than about 45%, more preferably more than about 50%, and most preferably more than about 60%. Microbial nucleotide sequences which have low GC contents may express poorly in plants due to the existence of ATTTA motifs which may destabilize messages, and AATAAA motifs which may cause inappropriate polyadenylation. Although preferred gene sequences may be adequately expressed in both monocotyledonous and dicotyledonous plant species, sequences can be modified to account for the specific codon preferences and GC content preferences of monocotyledons or dicotyledons as these preferences have been shown to differ (Murray et al. Nucl. Acids Res. 17: 477-498 (1989)). In addition, the nucleotide sequences are screened for the existence of illegitimate splice sites that may cause message truncation. All changes required to be made within the nucleotide sequences such as those described above are made using well known techniques of site directed mutagenesis, PCR, and synthetic gene construction using the methods described in the published patent applications EP 0 385 962 (to Monsanto), EP 0 359 472 (to Lubrizol, and WO 93/07278 (to Ciba-Geigy).

For efficient initiation of translation, sequences adjacent to the initiating methionine may require modification. For example, they can be modified by the inclusion of sequences known to be effective in plants. Joshi has suggested an appropriate consensus for plants

- 22 -

(NAR <u>15</u>: 6643-6653 (1987)) and Clontech suggests a further consensus translation initiator (1993/1994 catalog, page 210). These consensuses are suitable for use with the nucleotide sequences of this invention. The sequences are incorporated into constructions comprising the nucleotide sequences, up to and including the ATG (whilst leaving the second amino acid unmodified), or alternatively up to and including the GTC subsequent to the ATG (with the possibility of modifying the second amino acid of the transgene).

Expression of the nucleotide sequences in transgenic plants is driven by promoters shown to be functional in plants. The choice of promoter will vary depending on the temporal and spatial requirements for expression, and also depending on the target species. Thus, expression of the nucleotide sequences of this invention in leaves, in ears, in inflorescences (e.g. spikes, panicles, cobs, etc.), in roots, and/or seedlings is preferred. In many cases, however, protection against more than one type of insect pest is sought, and thus expression in multiple tissues is desirable. Although many promoters from dicotyledons have been shown to be operational in monocotyledons and vice versa, ideally dicotyledonous promoters are selected for expression in dicotyledons, and monocotyledonous promoters for expression in monocotyledons. However, there is no restriction to the provenance of selected promoters; it is sufficient that they are operational in driving the expression of the nucleotide sequences in the desired cell.

Preferred promoters that are expressed constitutively include promoters from genes encoding actin or ubiquitin and the CaMV 35S and 19S promoters. The nucleotide sequences of this invention can also be expressed under the regulation of promoters that are chemically regulated. This enables the insecticidal toxins to be synthesized only when the crop plants are treated with the inducing chemicals. Preferred technology for chemical induction of gene expression is detailed in the published application EP 0 332 104 (to Ciba-Geigy) and US patent 5,614,395. A preferred promoter for chemical induction is the tobacco PR-1a promoter.

A preferred category of promoters is that which is wound inducible. Numerous promoters have been described which are expressed at wound sites and also at the sites of phytopathogen infection. Ideally, such a promoter should only be active locally at the sites of infection, and in this way the insecticidal toxins only accumulate in cells which need to synthesize the insecticidal toxins to kill the invading insect pest. Preferred promoters of this kind include those described by Stanford *et al.* Mol. Gen. Genet. <u>215</u>: 200-208 (1989), Xu *et al.* Plant Molec. Biol. <u>22</u>: 573-588 (1993), Logemann *et al.* Plant Cell <u>1</u>: 151-158 (1989),

- 23 -

Rohrmeier & Lehle, Plant Molec. Biol. <u>22</u>: 783-792 (1993), Firek *et al.* Plant Molec. Biol. <u>22</u>: 129-142 (1993), and Warner *et al.* Plant J. <u>3</u>: 191-201 (1993).

Preferred tissue specific expression patterns include green tissue specific, root specific, stem specific, and flower specific. Promoters suitable for expression in green tissue include many which regulate genes involved in photosynthesis and many of these have been cloned from both monocotyledons and dicotyledons. A preferred promoter is the maize PEPC promoter from the phosphoenol carboxylase gene (Hudspeth & Grula, Plant Molec. Biol. 12: 579-589 (1989)). A preferred promoter for root specific expression is that described by de Framond (FEBS 290: 103-106 (1991); EP 0 452 269 to Ciba-Geigy). A preferred stem specific promoter is that described in US patent 5,625,136 (to Ciba-Geigy) and which drives expression of the maize *trpA* gene.

Especially preferred embodiments of the invention are transgenic plants expressing at least one of the nucleotide sequences of the invention in a root-preferred or root-specific fashion. Further preferred embodiments are transgenic plants expressing the nucleotide sequences in a wound-inducible or pathogen infection-inducible manner.

In addition to the selection of a suitable promoter, constructions for expression of an insecticidal toxin in plants require an appropriate transcription terminator to be attached downstream of the heterologous nucleotide sequence. Several such terminators are available and known in the art (e.g. tm1 from CaMV, E9 from rbcS). Any available terminator known to function in plants can be used in the context of this invention.

Numerous other sequences can be incorporated into expression cassettes described in this invention. These include sequences which have been shown to enhance expression such as intron sequences (e.g. from Adh1 and bronze1) and viral leader sequences (e.g. from TMV, MCMV and AMV).

It may be preferable to target expression of the nucleotide sequences of the present invention to different cellular localizations in the plant. In some cases, localization in the cytosol may be desirable, whereas in other cases, localization in some subcellular organelle may be preferred. Subcellular localization of transgene encoded enzymes is undertaken using techniques well known in the art. Typically, the DNA encoding the target peptide from a known organelle-targeted gene product is manipulated and fused upstream of the nucleotide sequence. Many such target sequences are known for the chloroplast and their functioning in heterologous constructions has been shown. The expression of the

- 24 -

nucleotide sequences of the present invention is also targeted to the endoplasmic reticulum or to the vacuoles of the host cells. Techniques to achieve this are well-known in the art.

Vectors suitable for plant transformation are described elsewhere in this specification. For *Agrobacterium*-mediated transformation, binary vectors or vectors carrying at least one T-DNA border sequence are suitable, whereas for direct gene transfer any vector is suitable and linear DNA containing only the construction of interest may be preferred. In the case of direct gene transfer, transformation with a single DNA species or co-transformation can be used (Schocher *et al.* Biotechnology 4: 1093-1096 (1986)). For both direct gene transfer and *Agrobacterium*-mediated transfer, transformation is usually (but not necessarily) undertaken with a selectable marker which may provide resistance to an antibiotic (kanamycin, hygromycin or methotrexate) or a herbicide (basta). The choice of selectable marker is not, however, critical to the invention.

In another preferred embodiment, a nucleotide sequence of the present invention is directly transformed into the plastid genome. A major advantage of plastid transformation is that plastids are generally capable of expressing bacterial genes without substantial modification, and plastids are capable of expressing multiple open reading frames under control of a single promoter. Plastid transformation technology is extensively described in U.S. Patent Nos. 5,451,513, 5,545,817, and 5,545,818, in PCT application no. WO 95/16783, and in McBride et al. (1994) Proc. Natl. Acad. Sci. USA 91, 7301-7305. The basic technique for chloroplast transformation involves introducing regions of cloned plastid DNA flanking a selectable marker together with the gene of interest into a suitable target tissue, e.g., using biolistics or protoplast transformation (e.g., calcium chloride or PEG mediated transformation). The 1 to 1.5 kb flanking regions, termed targeting sequences, facilitate homologous recombination with the plastid genome and thus allow the replacement or modification of specific regions of the plastome. Initially, point mutations in the chloroplast 16S rRNA and rps12 genes conferring resistance to spectinomycin and/or streptomycin are utilized as selectable markers for transformation (Svab, Z., Hajdukiewicz, P., and Maliga, P. (1990) Proc. Natl. Acad. Sci. USA 87, 8526-8530; Staub, J. M., and Maliga, P. (1992) Plant Cell 4, 39-45). This resulted in stable homoplasmic transformants at a frequency of approximately one per 100 bombardments of target leaves. The presence of cloning sites between these markers allowed creation of a plastid targeting vector for introduction of foreign genes (Staub, J.M., and Maliga, P. (1993) EMBO J. 12, 601-606). Substantial increases in transformation frequency are obtained by replacement of the recessive rRNA or r-protein antibiotic resistance genes with a dominant selectable marker, the bacterial

gene encoding the spectinomycin-detoxifying enzyme aminoglycoside-3'aadA adenyltransferase (Svab, Z., and Maliga, P. (1993) Proc. Natl. Acad. Sci. USA 90, 913-917). Previously, this marker had been used successfully for high-frequency transformation of the plastid genome of the green alga Chlamydomonas reinhardtii (Goldschmidt-Clermont, M. (1991) Nucl. Acids Res. 19: 4083-4089). Other selectable markers useful for plastid transformation are known in the art and encompassed within the scope of the invention. Typically, approximately 15-20 cell division cycles following transformation are required to reach a homoplastidic state. Plastid expression, in which genes are inserted by homologous recombination into all of the several thousand copies of the circular plastid genome present in each plant cell, takes advantage of the enormous copy number advantage over nuclearexpressed genes to permit expression levels that can readily exceed 10% of the total soluble plant protein. In a preferred embodiment, a nucleotide sequence of the present invention is inserted into a plastid targeting vector and transformed into the plastid genome of a desired plant host. Plants homoplastic for plastid genomes containing a nucleotide sequence of the present invention are obtained, and are preferentially capable of high expression of the nucleotide sequence.

Formulation of Insecticidal Compositions

The invention also includes compositions comprising at least one of the insecticidal toxins of the present invention. In order to effectively control insect pests such compositions preferably contain sufficient amounts of toxin. Such amounts vary depending on the crop to be protected, on the particular pest to be targeted, and on the environmental conditions, such as humidity, temperature or type of soil. In a preferred embodiment, compositions comprising the insecticidal toxins comprise host cells expressing the toxins without additional purification. In another preferred embodiment, the cells expressing the insecticidal toxins are lyophilized prior to their use as an insecticidal agent. In another embodiment, the insecticidal toxins are engineered to be secreted from the host cells. In cases where purification of the toxins from the host cells in which they are expressed is desired, various degrees of purification of the insecticidal toxins are reached.

The present invention further embraces the preparation of compositions comprising at least one insecticidal toxin of the present invention, which is homogeneously mixed with one or more compounds or groups of compounds described herein. The present invention also relates to methods of treating plants, which comprise application of the insecticidal toxins or compositions containing the insecticidal toxins, to plants. The insecticidal toxins

- 26 -

can be applied to the crop area in the form of compositions or plant to be treated, simultaneously or in succession, with further compounds. These compounds can be both fertilizers or micronutrient donors or other preparations that influence plant growth. They can also be selective herbicides, insecticides, fungicides, bactericides, nematicides, molluscicides or mixtures of several of these preparations, if desired together with further carriers, surfactants or application-promoting adjuvants customarily employed in the art of formulation. Suitable carriers and adjuvants can be solid or liquid and correspond to the substances ordinarily employed in formulation technology, e.g. natural or regenerated mineral substances, solvents, dispersants, wetting agents, tackifiers, binders or fertilizers.

A preferred method of applying insecticidal toxins of the present invention is by spraying to the environment hosting the insect pest like the soil, water, or foliage of plants. The number of applications and the rate of application depend on the type and intensity of infestation by the insect pest. The insecticidal toxins can also penetrate the plant through the roots via the soil (systemic action) by impregnating the locus of the plant with a liquid composition, or by applying the compounds in solid form to the soil, e.g. in granular form (soil application). The insecticidal toxins may also be applied to seeds (coating) by impregnating the seeds either with a liquid formulation containing insecticidal toxins, or coating them with a solid formulation. In special cases, further types of application are also possible, for example, selective treatment of the plant stems or buds. The insecticidal toxins can also be provided as bait located above or below the ground.

The insecticidal toxins are used in unmodified form or, preferably, together with the adjuvants conventionally employed in the art of formulation, and are therefore formulated in known manner to emulsifiable concentrates, coatable pastes, directly sprayable or dilutable solutions, dilute emulsions, wettable powders, soluble powders, dusts, granulates, and also encapsulations, for example, in polymer substances. Like the nature of the compositions, the methods of application, such as spraying, atomizing, dusting, scattering or pouring, are chosen in accordance with the intended objectives and the prevailing circumstances.

The formulations, compositions or preparations containing the insecticidal toxins and, where appropriate, a solid or liquid adjuvant, are prepared in known manner, for example by homogeneously mixing and/or grinding the insecticidal toxins with extenders, for example solvents, solid carriers and, where appropriate, surface-active compounds (surfactants).

Suitable solvents include aromatic hydrocarbons, preferably the fractions having 8 to 12 carbon atoms, for example, xylene mixtures or substituted naphthalenes, phthalates

such as dibutyl phthalate or dioctyl phthalate, aliphatic hydrocarbons such as cyclohexane or paraffins, alcohols and glycols and their ethers and esters, such as ethanol, ethylene glycol monomethyl or monoethyl ether, ketones such as cyclohexanone, strongly polar solvents such as N-methyl-2-pyrrolidone, dimethyl sulfoxide or dimethyl formamide, as well as epoxidized vegetable oils such as epoxidized coconut oil or soybean oil or water.

The solid carriers used e.g. for dusts and dispersible powders, are normally natural mineral fillers such as calcite, talcum, kaolin, montmorillonite or attapulgite. In order to improve the physical properties it is also possible to add highly dispersed silicic acid or highly dispersed absorbent polymers. Suitable granulated adsorptive carriers are porous types, for example pumice, broken brick, sepiolite or bentonite; and suitable nonsorbent carriers are materials such as calcite or sand. In addition, a great number of pregranulated materials of inorganic or organic nature can be used, e.g. especially dolomite or pulverized plant residues.

Suitable surface-active compounds are nonionic, cationic and/or anionic surfactants having good emulsifying, dispersing and wetting properties. The term "surfactants" will also be understood as comprising mixtures of surfactants. Suitable anionic surfactants can be both water-soluble soaps and water-soluble synthetic surface-active compounds.

Suitable soaps are the alkali metal salts, alkaline earth metal salts or unsubstituted or substituted ammonium salts of higher fatty acids (chains of 10 to 22 carbon atoms), for example the sodium or potassium salts of oleic or stearic acid, or of natural fatty acid mixtures which can be obtained for example from coconut oil or tallow oil. The fatty acid methyltaurin salts may also be used.

More frequently, however, so-called synthetic surfactants are used, especially fatty sulfonates, fatty sulfates, sulfonated benzimidazole derivatives or alkylarylsulfonates.

The fatty sulfonates or sulfates are usually in the form of alkali metal salts, alkaline earth metal salts or unsubstituted or substituted ammonium salts and have a 8 to 22 carbon alkyl radical which also includes the alkyl moiety of alkyl radicals, for example, the sodium or calcium salt of lignonsulfonic acid, of dodecylsulfate or of a mixture of fatty alcohol sulfates obtained from natural fatty acids. These compounds also comprise the salts of sulfuric acid esters and sulfonic acids of fatty alcohol/ethylene oxide adducts. sulfonated benzimidazole derivatives preferably contain 2 sulfonic acid groups and one fatty acid radical containing 8 to 22 carbon atoms. Examples of alkylarylsulfonates are the sodium, calcium triethanolamine salts dodecylbenzenesulfonic or of dibutylnapthalenesulfonic acid, or of naphthalenesulfonic а acid/formaldehyde

condensation product. Also suitable are corresponding phosphates, e.g. salts of the phosphoric acid ester of an adduct of p-nonylphenol with 4 to 14 moles of ethylene oxide.

Non-ionic surfactants are preferably polyglycol ether derivatives of aliphatic or cycloaliphatic alcohols, or saturated or unsaturated fatty acids and alkylphenols, said derivatives containing 3 to 30 glycol ether groups and 8 to 20 carbon atoms in the (aliphatic) hydrocarbon moiety and 6 to 18 carbon atoms in the alkyl moiety of the alkylphenols.

Further suitable non-ionic surfactants are the water-soluble adducts of polyethylene oxide with polypropylene glycol, ethylenediamine propylene glycol and alkylpolypropylene glycol containing 1 to 10 carbon atoms in the alkyl chain, which adducts contain 20 to 250 ethylene glycol ether groups and 10 to 100 propylene glycol ether groups. These compounds usually contain 1 to 5 ethylene glycol units per propylene glycol unit.

Representative examples of non-ionic are nonylphenolpolyethoxyethanols, castor oil polyglycol ethers, polypropylene/polyethylene oxide adducts, tributylphenoxypolyethoxyethanol, polyethylene glycol and octylphenoxyethoxyethanol. Fatty acid esters of polyoxyethylene sorbitan and polyoxyethylene sorbitan trioleate are also suitable non-ionic surfactants.

Cationic surfactants are preferably quaternary ammonium salts which have, as N-substituent, at least one C8-C22 alkyl radical and, as further substituents, lower unsubstituted or halogenated alkyl, benzyl or lower hydroxyalkyl radicals. The salts are preferably in the form of halides, methylsulfates or ethylsulfates, e.g. stearyltrimethylammonium chloride or benzyldi(2-chloroethyl)ethylammonium bromide.

The surfactants customarily employed in the art of formulation are described, for example, in "McCutcheon's Detergents and Emulsifiers Annual," MC Publishing Corp. Ringwood, New Jersey, 1979, and Sisely and Wood, "Encyclopedia of Surface Active Agents," Chemical Publishing Co., Inc. New York, 1980.

EXAMPLES

The invention will be further described by reference to the following detailed examples. These examples are provided for purposes of illustration only, and are not intended to be limiting unless otherwise specified. Standard recombinant DNA and molecular cloning techniques used here are well known in the art and are described by Ausubel (ed.), Current Protocols in Molecular Biology, John Wiley and Sons, Inc. (1994); T. Maniatis, E. F. Fritsch and J. Sambrook, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor laboratory, Cold Spring Harbor, NY (1989); and by T.J. Silhavy, M.L. Berman, and L.W. Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984).

A. Isolation Of Nucleotide Sequences Whose Expression Results In Toxins Active Against Lepidopteran Insects

Example 1: Construction of Cosmid Library from Photorhabdus luminescens

Photorhabdus luminescens strain ATCC 29999 is grown in nutrient broth at 25°C for three days as described in the ATCC protocol for bioassay. The culture is grown for 24 hours for DNA isolation. Total DNA is isolated by treating freshly grown cells resuspended in 100 mM Tris pH 8, 10 mM EDTA with 2 mg/ml lysozyme for 30 minutes at 37°C. Proteinase K is added to a final concentration of 100 mg/ml, SDS is added to a final concentration of 0.5% SDS and the sample is incubated at 45°C. After the solution becomes clear and viscous, the SDS concentration is raised to 1%, and 300 mM NaCl and an equal volume of phenol-chloroform-isoamyl alcohol are added, mixed gently for 5 minutes and centrifuged at 3K. The phenol-chloroform-isoamyl alcohol extraction is repeated twice. The aqueous phase is mixed with 0.7 volumes isopropanol, and the sample is centrifuged. The pellet is washed 3 times with 70% ethanol and the nucleic acids are gently resuspended in 0.5X TE.

The DNA is treated with 0.3 units of Sau3A per mg DNA at 37°C for 3.5 minutes in 100 ml volume containing a total of 6 mg DNA. The reaction is then heated for 30 minutes at 65°C to inactivate the enzyme. Then 2 units of Calf Intestinal Alkaline Phosphatase are added and incubated for 30 minutes at 37°C. The sample is mixed with an equal volume of

phenol-chloroform-isoamyl alcohol and centrifuged. The aqueous phase is removed, precipitated with 0.7 volume isopropanol and centrifuged. The supernatant is transferred to a fresh tube, precipitated with ethanol, and the nucleic acids are resuspended in 0.5X TE at a concentration of 100 hg/ml.

SuperCos cosmid vector (Stratagene, La Jolla, CA) is prepared as described by the supplier utilizing the *BamHI* cloning site. Prepared SuperCos at 100 hg/ml is ligated with the *Sau3A* digested *P.luminescens* DNA at a molar ratio of 2:1 in a 5 ml volume overnight at 6°C. The ligation mixture is packaged using Gigapack XL III (Stratagene), as described by the supplier. Packaged phages are used to infect XL-1MR (Stratagene) cells as described by the supplier. The cosmid library is plated on L-agar with 50 mg/ml kanamycin and incubated 16 hours at 37°C. 500 colonies are patched onto fresh L-kan plates at 50 colonies per plate. From the other plates the cells are washed off with L broth and mixed with 20% glycerol and frozen at -80°C.

Example 2: Insect Bioassays

Plutella xylostella bioassays are performed by aliquoting of 50 μl of the *E. coli* culture on the solid artificial *Plutella xylostella* diet (Biever and Boldt, *Annals of Entomological Society of America*, 1971; Shelton et al., *J. Ent. Sci.* 26:17). 4 ml of the diet is poured into 1 oz. clear plastic cups (Bioserve product #9051). 5 neonate *P. xylostella* from a diet adapted lab colony are placed in each diet-containing cup and then covered with a white paper lid (Bioserve product #9049). 10 larvae are assayed per concentration. Trays of cups are placed in an incubator for 3 days at 72°F with a 14:10 (hours) light:dark cycle. Then, the number of live larvae in each cup is recorded. Bioassays for other insects are performed as described for *Plutella xylostella*, but using the diet required by the insect to be tested.

The broth of *P. luminescens* undiluted and diluted 1:100 gives 100% mortality against *P. xylostella*. The broth of *P. luminescens* also gives 100% mortality against *Diabrotica virgifera virgifera*. Three clones with activity against *P. xylostella* and *Heliothis virescens* are obtained after screening 500 *E. coli* clones by insect bioassay. These cosmid clones are given the numbers pCIB9349, pCIB9350, and pCIB9351.

- 31 -

Example 3: Isolation of the Nucl otid Sequence Responsible for Insect C ntrol Activity from Clones pClB9349, pClB9350, and pClB9351

The three clones pClB9349, pClB9350 and pClB9351 are found to be overlapping cosmids by restriction enzyme mapping. After digestion with PacI, clones pCIB9349 and pCIB9351 give two DNA fragments each, and pCIB9350 gives three DNA fragments. Each fragment is isolated and is self-ligated. The enzyme Pacl does not cut the SuperCos vector; therefore, only fragments linked to it are re-isolated. The ligation mixtures are transformed into DH5α E. coli cells. Isolated transformed bacterial colonies are grown in L broth with 50 μg/ml kanamycin, and plasmid DNA is isolated by using the alkaline miniprep protocol as described in Sambrook, et al. DNA is digested with Notl/Pacl and two clones, pCIB9355 and pCIB9356, are found by bioassay to still contain the insecticidal activity. Clone pCIB9355 is digested with Not and a 17 kb and a 4 kb DNA fragment are generated. The 17 kb fragment is isolated and ligated into Bluescript vector previously cut with Notl and transformed into DH5\alpha E. coli cells. The isolated transformed bacterial colonies are grown as described and plasmid DNA is isolated by the alkaline miniprep protocol. A clone containing the 17 kb insert is named pCIB9359 and tested by bioassay. The results are shown in Example 5. 3 µg of the 17 kb insert is isolated and treated with 0.3 unit of Sau3A per µg DNA for 4, 6, and 8 minutes at 37°C, heated at 75°C for 15 minutes. The samples are pooled and ligated into pUC19 previously cut with BamHI and treated with calf intestinal alkaline phosphatase. The ligation is transformed into DH5lpha cells and plated on L agar with Xgal/Amp as described in Sambrook et al. and grown overnight at 37°C. White colonies are picked and grown in L broth with 100 µg/ml and plasmid DNA is isolated as previously described. DNA is digested with *EcoRI/HindIII* and novel restriction patterns are sequenced. Sequencing primers are ordered from Genosys Biotechnologies (Woodlands, TX). Sequencing is performed using the dideoxy chain-termination method. Sequencing is completed using Applied Biosystems Inc. model 377 automated DNA sequencer (Foster City, CA). Sequence is assembled using 3.0 from Gene Codes Corporation (Ann Arbor, MI).

Exampl 4: Subcloning f th 9.7 kb EcoRI/Xbal Fragment Fr m pCIB9359

pClB9359 is digested with *EcoRI* and *XbaI* and the DNA is run on a 0.8% Seaplaque/TBE gel. The 9.7 kb fragment (SEQ ID NO:1) is isolated and ligated into pUC19 previously digested with *EcoRI* and *XbaI*. The ligation mixture is transformed into DH5 α *E. coli* cells. Transformed bacteria are grown and plasmid DNA is isolated as previously described. The vector containing the 9.7 kb fragment in pUC19 is designated pClB9359-7 and bioassay results are shown in Example 5.

Example 5: Bioassay Results for Cosmid Clones pCIB9359 and pCIB9359-7

Cultures of *E. coli* strains 9359 and 9359-7 containing clones pCIB9359 and pCIB9359-7, respectively, are tested for insecticidal activity against the following insects in insect bioassays:

Insects	Clones
	pClB9359 and pClB9359-7
Plutella xylostella (Diamondback Moth (DBM))	+++
Heliothis virescens (Tobacco Budworm (TBW))	++
Helicoverpa zea (Corn Earworm (CEW))	+++
Spodoptera exigua (Beet Armyworm (BAW))	+
Spodoptera frugiperda (Fall Armyworm (FAW))	+
Trichoplusia ni (Cabbage Looper (CL))	+++
Ostrinia nubilalis (European Corn Borer (ECB))	++
Manduca sexta (Tobacco Hornworm (THW)	na
Diabrotica virgifera (Western Corn Rootworm (WCR))	na
Agrotis ipsilon (Black Cutworm (BCW))	na

na = not active

- + = significant growth inhibition
- ++ = >40% mortality, but less than 100%
- +++ = 100% mortality

The clones show insecticidal activity against *P. xylostella*, *H. virescens*, *H. zea*, *T. ni*, and *O. nubilalis*, and significant insect control activity against *S. exigua* and *S. frugiperda*.

Example 6: Identification of Active Region of pClB9359-7 By Subcloning

Cultures of *E. coli* strains containing subclones of pClB9359-7 are tested for insecticidal activity in insect bioassays against *P. xylostella*.

Restriction	Nucleotide Posit	ion Relative to 9.7 kb	Insecticidal Activity Against	
Fragment	EcoRI/Xbal fragn	nent (SEQ ID NO:1)	Plutella xylostella	
	from pCIB9539-7	and Size in kb		
EcoRI/XbaI	1 to 9712	9.7 kb	+++	
EcoRV	(-912) to 2309	3.2 kb	na	
HindIII	665 to 5438	4.7 kb	na	
Kpnl	1441 to 8137	6.9 kb	na	
Sacl/Xbal	2677 to 9712	7.0 kb	na	

na = not active

Example 7: Characterization of pCIB9359-7 Insect Control Activity By Titration

Dilutions of a culture of E.coli strain 9359-7 containing pCIB9359-7 are tested for insecticidal activity in insect bioassays. Dilutions are prepared in a culture of E.coli XL-1 in a total volume of 100 μ l and are transferred to diet cups with 5 insects per cup. The results show the percentage (%) of insect mortality.

^{+ =} significant growth inhibition

^{++ = &}gt;40% mortality, but less than 100%

^{+++ = 100%} mortality

- 34 -

μ l 9359-7 Culture	Px	Hv	Hz	Tn
100	100	72	48	100
50	100	84	68	92
25	100	52	32	100
12.5	96	52	36	68
6.25	88	20	4	32
0	36	20	24	0

Px = P. xylostella, Hv = H. virescens, Hz = H. zea, Tn = T. ni.

Cultures of E. coli 9359-7 still show substantial insecticidal activity after dilution.

Example 8: Stability of pCIB9359-7 Activity

The stability of the toxins is tested after storage for 2 weeks at different temperatures and conditions. 300 ml of Luria broth containing 100 (μg/ml ampicillin is inoculated with *E. coli* strain 9359-7 and grown overnight at 37°C. Samples are placed in sterile 15 ml screw cap tubes and stored at 22°C and 4°C. Another sample is centrifuged; the supernatant is removed, freeze dried and stored at 22°C. The samples are stored under these conditions for 2 weeks and then a bioassay is conducted against *P. xylostella*. The freeze dried material is resuspended in the same volume as before. All samples are resuspended by vortexing.

Conditions	Results
22°C (2 weeks)	+++
4°C (2 weeks)	+++
Freeze Dried (2 weeks)	+++

na = not active; + = significant growth inhibition; ++ = >40% mortality, but less than 100%; +++ = 100% mortality

WO 99/42589 PCT/EP99/01015

- 35 -

This demonstrates that the toxins retain their activity for at least two weeks at 22°C, 4°C, and freeze-dried, and are therefore very stable.

Example 9: Size Fraction of pCIB9359-7 Activity

The approximate sizes of the insecticidal toxins are determined. P. luminescens cosmid clones pClB9359-7 and pUC19 in E. coli host DH5α are grown in media consisting of 50% Terrific broth and 50% Luria broth, supplemented with 50 μg/ml ampicillin. Cultures (three tubes of each strain) are inoculated into 3 ml of the above media in culture tubes and incubated on a roller wheel overnight at 37°C. Cultures of each strain are combined and sonicated using a Branson Model 450 Sonicator, micro tip, for approximately six 10 second cycles with cooling on ice between cycles. The sonicates are centrifuged in a Sorvall SS34 rotor at 6000 RPM for 10 minutes. The resultant supernatants are filtered through a 0.2 µ filter. The 3 ml fractions of the filtrates are applied to Bio-Rad Econo-Pac 10DG columns that have been previously equilibrated with 10 ml of 50mM NaCl, 25 mM Tris base, pH 7.0. The flow through collected during sample loading is discarded. The samples are fractionated with two subsequent additions of 4 ml each of the NaCl - Tris equilibration buffer. The two four ml fractions are saved for testing. The first fraction contains all material above about 6,000 mol. wt; the second fraction contains material smaller than 6,000 mol. wt. A sample of the whole culture broth, the sonicate, and the filtered supernatant on the sonicate are tested along with the three fractions from the 10DG column for activity on P. xylostella neonates in bioassays.

The culture, the sonicate, and the filtered supernatant of the sonicate, and the first column fraction from the 9359-7 sample are highly active on *P. xylostella*. The second column fraction from 9359-7 is slightly active (some stunting only). No activity is found in the third fraction from 9359-7. The sample from DH5-pUC19 does not have any activity. This indicates that the molecular weights of the toxins are above 6,000.

Example 10: Heat Inactivitation of pCIB9359-7 Activity

The heat stability of the toxins is determined. Overnight cultures of the *E. coli* strain pClB9359-7 are grown in a 50:50 mixture of Luria broth and Terrific broth. Cultures are grown at 37°C in culture tubes on a tube roller. A one ml sample of the culture is placed in

a 1.5 ml eppendorf tube and placed in a boiling water bath. The sample is removed after five minutes and allowed to cool to room temperature. This sample along with an untreated portion of the culture is assayed on *P. xylostella*. 50µl of sample of sample is spread on diet, allowed to dry and neonate larvae *P. xylostella* applied to the surface. The assay is incubated for 5 days at room temperature.

The untreated sample causes 100% mortality. The heat treated sample and a diet alone control do not cause any observable mortality, showing the toxins are heat sensitive.

Example 11: Leaf Dip Bioassay of pCIB9359-7

Insecticidal activity of the toxins is tested in a leaf dip bioassay. Six leaves approximately 2cm in diameter each are cut from seedlings of turnip and placed in a 1oz. plastic cup (Jet Plastica) with 4ml-5ml of the resuspended toxin, covered tightly, and shaken until thoroughly wetted. The treated leaves are placed in 50mm petri dishes (Gelman Sciences) on absorbent pads moistened with 300µl of water. The dish covers are left open until the leaf surface appears dry and then placed on tightly so that the leaves do not dry out.

Ten neonate *P. xylostella* larvae are placed in each petri dish arena. Also, a treatment of 0.1% Bond spreader/sticker with no toxin is set up as a control. The arenas are monitored daily for signs of drying leaves, and water is added or leaves replaced if necessary. After 3 days the leaves and arenas are examined under a dissecting microscope, and the number of live larvae in each arena is recorded.

100% mortality is found for 9359-7 and none in the no-toxin control, showing that the toxins are also insecticidal in a leaf dip assay.

B. Isolation Of Nucleic Acid Sequences Whose Expression Results In Toxins Active Against Lepidopteran and Coleopteran Insects

Example 12: Total DNA Isolation from Photorhabdus luminescens

Photorhabdus luminescens strain ATCC 29999 is grown 14-18 hours in L broth. Total DNA is isolated from 1.5 mls of culture resuspended in 0.5% SDS, 100μg/ml proteinase K, TE to a final volume of 600 μl. After a 1 hour incubation at 37°C, 100μl 5M

NaCl and 80μ l CTAB/NaCl are added and the culture is incubated at 65° C for 10 minutes. An equal volume of chloroform is added; the culture is mixed gently and spun. The aqueous phase is extracted once with phenol and once with chloroform. The nucleic acids are treated with 10 μ g RNase A for 30 minutes at room temperature. The aqueous phase is mixed with 0.6 volumes isopropanol and the sample is centrifuged. The pellet is washed once with 70% ethanol and the nucleic acids are gently resuspended in 100-200ul TE.

Example 13: PCR Amplification of Probes

Two probes are PCR amplified from *Photorhabdus luminescens* strain ATCC 29999 genomic DNA using oligos 5'-ACACAGCAGGTTCGTCAG-3' (SEQ ID NO:7) and 5'-GGCAGAAGCACTCAACTC-3' (SEQ ID NO:8) to amplify probe #1 and oligos 5'-ATTGATAGCACGCGGCGACC-3' (SEQ ID NO:9) and 5'-

TTGTAACGTGGAGCCGAACTGG-3' (SEQ ID NO:10) to amplify probe #2. The oligos are ordered from Genosys Biotechnologies, Inc. (Texas). Approximately 10-50 ng of genomic DNA is used as the template. 0.8µM of oligos, 200µM of dNTPs, 1X Taq DNA Polymerase buffer and 2.5 units of Taq DNA Polymerase are included in the reaction. The reaction conditions are as follows:

94°C - 1 minute

94°C - 30 seconds / 60°C - 30 seconds / 72°C - 30 seconds (25 cycles)

72°C - 5 minutes

4°C - indefinite soak

The reactions are preferably carried out in a PCR System 9600 (Perkin Elmer) thermocycler.

Example 14: Probing a Photorhabdus luminescens Library

600 clones from the *P. luminescens* cosmid library described in Example 1 are patched to L-amp plates in duplicate. The colonies are grown overnight then moved to 4°C. The colonies are lifted onto Colony/Plaque Screen Hybridization Transfer Membranes (Biotechnology Systems NEN Research Products). The membranes are incubated 2-3 minutes in 0.75ml 0.5N NaOH twice. The membranes are then incubated 2-3 minutes in

0.75ml 1.0M Tris-HCl, pH 7.5 twice. The membranes are allowed to dry at room temperature.

Probe #1 and probe #2 described in Example 13 are labeled using the DECAprime II Kit as described by the manufacturer (Ambion cat# 1455). Unincorporated nucleotides are removed from the labeled probes using Quick Spin Columns as described by the manufacturer (Boehringer Mannheim cat #1273973). The labeled probes are measured for incorporated radioactivity and the specific activity is 10,000,000 cpm. Membranes are prewetted with 2X SSC and hybridized with the probes for 12-16 hours at 65°C. One set of colony lifts is hybridized with probe #1 and the other set is hybridized with probe #2. The membranes are washed with wash CHURCH solutions 1 and 2 (Church and Gilbert, *Proc. Natl. Acad. Sci. USA* 81:1991-1995 (1984)) and exposed to Kodak film.

Twenty one clones are identified that hybridize to probe #1 and seven clones are identified that hybridize to probe #2. The gene in the clones isolated with probe #1 is named *hph1* and the gene in the clones isolated with probe #2 is named *hph2*.

Example 15: Insect Bioassays

The clones identified in Example 14 are tested for insecticidal activity against the following insects in insect bioassays: *Diabrotica virgifera virgifera* (Western Corn Rootworm (WCR)), *Diabrotica undecimpunctata howardi* (Southern Corn Rootworm (SCR)), *Ostrinia nubilalis* (European Corn Borer (ECB)), and *Plutella xylostella* (Diamondback Moth (DBM)).

Diabrotica virgifera virgifera (Western Corn Rootworm) and Diabrotica undecimpunctata howardi (Southern Corn Rootworm) assays are performed using a diet incorporation method. 500μl of an overnight culture of the cosmid library in XL-1 Blue MR cells (Stratagene) is sonicated and then mixed with 500μl of diet. Once the diet solidifies, it is dispensed in a petri dish and 20 larvae are introduced over the diet. Trays of dishes are placed in an incubator for 3-5 days, and percent mortality is recorded at the end of the assay period.

Ostrinia nubilalis (European Corn Borer) and Plutella xylostella (Diamondback Moth) assays are performed by a surface treatment method. The diet is poured in the petri dish and allowed it to solidify. The E. coli culture of 200 -300µl volume is dispensed over the diet surface and entire diet surface is covered to spread the culture with the help of bacterial loop. Once the surface is dry, 10 larvae are introduced over the diet surface. Trays of

dishes are placed in an incubator for 3-5 days. The assay with European Corn Borer is incubated at 30°C in complete darkness; the assay with Diamondback Moth is incubated at 72°F with a 14:10 (hours) light:dark cycle. Percent mortality is recorded at the end of the assay period.

Cosmids containing *hph2* are identified with a range of activities, including: WCR only; SCR only; WCR and SCR; SCR and ECB; WCR, SCR, and ECB; or WCR, SCR, ECB, and DBM activity.

In addition to probing the *P. luminescens* cosmid library with DNA probes, 600 clones are screened by Western Corn Rootworm bioassay. A clone is identified with activity against Western Corn Rootworm. This clone hybridizes with probe #2.

From these bioassays, cosmid 514, having activity against WCR, SCR, ECB, and DBM, is selected for sequencing.

Example 16: Sequencing of Cosmid 514

Cosmid 514 is sequenced using dye terminator chemistry on an ABI 377 instrument. The nucleotide sequence of cosmid 514 is set forth as SEQ ID NO:11. Cosmid 514 is designated pNOV2400 and deposited with the NRRL in $E.\ coli\ DH5\alpha$ and assigned accession no. B-30077.

Example 17: Subcloning Insecticidal Regions of Cosmid 514

514a

An 9011 base pair fragment within cosmid 514 (SEQ ID NO:11) is removed by digesting the cosmid with the restriction endonuclease *Spel* (New England Biolabs (Massachusetts), and ligating (T4 DNA Ligase, NEB) the remainder of 514. Subclone 514a consists of cosmid 514 DNA from base pairs 1-2157 ligated to base pairs 11,169-37,948.

H2O2/pET34

hph2 and orf2 (SEQ ID NO:11, base pairs 23,768-35,838) are cloned into pET34b (Novagen, Wisconsin). Restriction sites are engineered on both ends of each gene to facilitate cloning. PCR is used to add the restriction sites to the genes. A BamHI site is on the 5' end of hph2 immediately upstream of the ATG of hph2, and a Sac site is added to

- 40 -

the 3' end of hph2 immediately following the DNA triplet encoding the stop codon. A guanidine is added between the BamHI site and the start codon of hph2 to put the hph2 gene in frame with the Cellulose Binding Domain tag in pET34b. Orf2 has a SacI site upstream of the 56 base pairs between the stop codon of hph2 and the start codon of orf2. The 56 base pairs are included in the hph2-orf2 construct to mimic their setup in the 514 cosmid. Orf2 has an XhoI site on the 3' end immediately following the stop codon. The oligos used to add the restriction sites to hph2 and orf2 are as follows:

hph2-A	5'-CGGGATCCGATGATTTTAAAAGG-3' (SEQ ID NO:15)
hph2-B	5'-GCGCCATTGATTTGAG-3' (SEQ ID NO:16)
hph2-C	5'-CATTAGAGGTCGAACGTAC-3' (SEQ ID NO:17)
hph2-D	5'-GAGCGAGCTCTTACTTAATGGTGTAG-3' (SEQ ID NO:18)
orf2-A3	5'-CAGCGAGCTCCATGCAGAATTCACAGAC-3' (SEQ ID NO:19)
orf2-B	5'-GGCAATGGCAGCGATAAG-3' (SEQ ID NO:20)
orf2-C	5'-CATTAACGCAGGAAGAGC-3' (SEQ ID NO:21)
orf2-D	5'-GACCTCGAGTTACACGAGCGCGTCAG-3' (SEQ ID NO:22)

The BamHI-Sacl 7583 base pair fragment, corresponding to the hph2 gene, and the Sacl-Xhol 4502 base pair orf2 (including the 56 base pairs between hph2 and orf2 open reading frames), corresponding to orf2, are ligated with BamHI-Xhol-digested vector DNA pET34b.

Orf5/pBS (Noti-BamHI)

The 5325 base pair *Not*I-*Bam*HI fragment of cosmid 514 is cloned into pBS-SK using *AfI*III-*Not*I (415 bp) and *Bam*HI-*AfI*III (2530 bp) fragments of pBS-SK.

O5-H2-O2

The 12,031 base pair *BamHI-XhoI* fragment of H2O2/pET34 is cloned into the 8220 base pair *XhoI-BamHI* fragment of Orf5/pBS.

O51011H2O2

A 7298 base pair *Bam*HI-*Mlu*I fragment from subclone 514a is ligated (T4 DNA Ligase, NEB) with 9588 bp *Mlu*I-*Xho*I and 8220 bp *Xho*I-*Bam*HI fragments of subclone O5-H2-O2. The resulting ~ 22 kb subclone O51011H2O2, which has activity against WCR and

WO 99/42589 PCT/EP99/01015

ECB, is designated pNOV1001 and deposited with the NRRL in *E. coli* DH5 α and assigned accession no. B-30078.

- 41 -

AKH2O2

A 12,074 base pair *BamHI-AvrII* fragment of H2O2/pET34 is ligated (T4 DNA Ligase, NEB) into pK184 *NheI-BamHI* fragment (2228 bp), generating a clone containing hph2 and orf2 in a p15a origin of replication, kanamycin-resistant vector.

Example 18: Insecticidal Activity of Subclones

Bioassays as described above are performed with *E. coli* cultures that express the above subclones, both singly and in combination. Coexpressing AKH2O2 and Orf5/pBS in *E. coli*, for example in DH5α or HB101, is found to give insecticidal activity against the Lepidopterans *Plutella xylostella* (Diamondback Moth), *Ostrinia nubilalis* (European Corn Borer), and *Manduca sexta* (Tobacco Hornworm), as well as against the Coleopterans *Diabrotica virgifera virgifera* (Western Corn Rootworm), *Diabrotica undecimpunctata howardi* (Southern Corn Rootworm), and *Leptinotarsa decimlineata* (Colorado Potato Beetle). Thus, coexpression of hph2 (SEQ ID NO:11, base pairs 23,768-31,336), orf2 (SEQ ID NO:11, base pairs 31,393-35,838), and orf5 (SEQ ID NO:11, base pairs 15,171-18,035) is sufficient to control these insects. In addition, expression of each of these three ORFs on separate plasmids gives insect control activity, demonstrating that they do not have to be genetically linked to be active, so long as all three gene products are present.

C. Expression of the Nucleic Acid Sequences of the Invention in Heterologous Microbial Hosts

Microorganisms which are suitable for the heterologous expression of the nucleotide sequences of the invention are all microorganisms which are capable of colonizing plants or the rhizosphere. As such they will be brought into contact with insect pests. These include gram-negative microorganisms such as *Pseudomonas, Enterobacter* and *Serratia*, the gram-positive microorganism *Bacillus* and the fungi *Trichoderma, Gliocladium*, and *Saccharomyces cerevisiae*. Particularly preferred heterologous hosts are *Pseudomonas fluorescens, Pseudomonas putida, Pseudomonas cepacia, Pseudomonas aureofaciens,*

Pseudomonas aurantiaca, Enterobacter cloacae, Serratia marscesens, Bacillus subtilis, Bacillus cereus, Trichoderma viride, Trichoderma harzianum, Gliocladium virens, and Saccharomyces cerevisiae.

Example 19: Expression of the Nucleotide Sequences in *E. coli* and Other Gram-Negative Bacteria

Many genes have been expressed in gram-negative bacteria in a heterologous manner. Expression vector pKK223-3 (Pharmacia catalogue # 27-4935-01) allows expression in *E. coli*. This vector has a strong *tac* promoter (Brosius, J. *et al.*, *Proc. Natl. Acad. Sci. USA 81*) regulated by the *lac* repressor and induced by IPTG. A number of other expression systems have been developed for use in *E. coli*. The thermoinducible expression vector pPL (Pharmacia #27-4946-01) uses a tightly regulated bacteriophage λ promoter which allows for high level expression of proteins. The *lac* promoter provides another means of expression but the promoter is not expressed at such high levels as the *tac* promoter. With the addition of broad host range replicons to some of these expression system vectors, expression of the nucleotide sequence in closely related gram negative-bacteria such as *Pseudomonas*, *Enterobacter*, *Serratia* and *Erwinia* is possible. For example, pLRKD211 (Kaiser & Kroos, Proc. Natl. Acad. Sci. USA <u>81</u>: 5816-5820 (1984)) contains the broad host range replicon *ori T* which allows replication in many gram-negative bacteria.

In *E. coli*, induction by IPTG is required for expression of the *tac* (*i.e. trp-lac*) promoter. When this same promoter (*e.g.* on wide-host range plasmid pLRKD211) is introduced into *Pseudomonas* it is constitutively active without induction by IPTG. This *trp-lac* promoter can be placed in front of any gene or operon of interest for expression in *Pseudomonas* or any other closely related bacterium for the purposes of the constitutive expression of such a gene. Thus, a nucleotide sequence whose expression results in an insecticidal toxin can therefore be placed behind a strong constitutive promoter, transferred to a bacterium which has plant or rhizosphere colonizing properties turning this organism to an insecticidal agent. Other possible promoters can be used for the constitutive expression of the nucleotide sequence in gram-negative bacteria. These include, for example, the promoter from the *Pseudomonas* regulatory genes *gafA* and *lemA* (WO 94/01561) and the

Pseudomonas savastanoi IAA operon promoter (Gaffney et al., J. Bacteriol. 172: 5593-5601 (1990).

Example 20: Expression of the Nucleotide Sequences in Gram-Positive Bacteria

Heterologous expression of the nucleotides sequence in gram-positive bacteria is another means of producing the insecticidal toxins. Expression systems for *Bacillus* and *Streptomyces* are the best characterized. The promoter for the erythromycin resistance gene (*ermR*) from *Streptococcus pneumoniae* has been shown to be active in gram-positive aerobes and anaerobes and also in *E.coli* (Trieu-Cuot *et al.*, Nucl Acids Res 18: 3660 (1990)). A further antibiotic resistance promoter from the thiostreptone gene has been used in *Streptomyces* cloning vectors (Bibb, Mol Gen Genet 199: 26-36 (1985)). The shuttle vector pHT3101 is also appropriate for expression in *Bacillus* (Lereclus, FEMS Microbiol Lett 60: 211-218 (1989)). A significant advantage of this approach is that many grampositive bacteria produce spores which can be used in formulations that produce insecticidal agents with a longer shelf life. *Bacillus* and *Streptomyces* species are aggressive colonizers of soils

Example 21: Expression of the Nucleotide Sequences in Fungi

Trichoderma harzianum and Gliocladium virens have been shown to provide varying levels of biocontrol in the field (US 5,165,928 and US 4,996,157, both to Cornell Research Foundation). A nucleotide sequence whose expression results in an insecticidal toxin could be expressed in such a fungus. This could be accomplished by a number of ways which are well known in the art. One is protoplast-mediated transformation of the fungus by PEG or electroporation-mediated techniques. Alternatively, particle bombardment can be used to transform protoplasts or other fungal cells with the ability to develop into regenerated mature structures. The vector pAN7-1, originally developed for Aspergillus transformation and now used widely for fungal transformation (Curragh et al., Mycol. Res. 97(3): 313-317 (1992); Tooley et al., Curr. Genet. 21: 55-60 (1992); Punt et al., Gene 56: 117-124 (1987)) is engineered to contain the nucleotide sequence. This plasmid contains the E. coli the hygromycin B resistance gene flanked by the Aspergillus nidulans gpd promoter and the trpC terminator (Punt et al., Gene 56: 117-124 (1987)).

- 44 -

In a preferred embodiment, the nucleic acid sequences of the invention are expressed in the yeast *Saccharomyces cerevisiae*. Each of the three ORF's of SEQ ID NO:11 (hph2, orf2 and orf5), which together confer insecticidal activity, are cloned into individual vectors with the GAL1 inducible promoter and the CYC1 terminator. Each vector has ampicillin resistance and the 2 micron replicon. The vectors differ in their yeast growth markers. hph2 is cloned into p424 (TRP1, ATCC 87329), orf2 into p423 (HIS3, ATCC 87327), and orf5 into p425 (LEU2, ATCC 87331). The three constructs are transformed into *S. cerevisiae* independently and together. The three ORFs are expressed together and tested for protein expression and insecticidal activity.

D. Expression of the Nucleotide Sequences in Transgenic Plants

The nucleic acid sequences described in this application can be incorporated into plant cells using conventional recombinant DNA technology. Generally, this involves inserting a coding sequence of the invention into an expression system to which the coding sequence is heterologous (i.e., not normally present) using standard cloning procedures known in the art. The vector contains the necessary elements for the transcription and translation of the inserted protein-coding sequences. A large number of vector systems known in the art can be used, such as plasmids, bacteriophage viruses and other modified viruses. Suitable vectors include, but are not limited to, viral vectors such as lambda vector systems λqtl1, λqtl0 and Charon 4; plasmid vectors such as pBl121, pBR322, pACYC177, pACYC184, pAR series, pKK223-3, pUC8, pUC9, pUC18, pUC19, pLG339, pRK290, pKC37, pKC101, pCDNAII; and other similar systems. The components of the expression system may also be modified to increase expression. For example, truncated sequences, nucleotide substitutions or other modifications may be employed. The expression systems described herein can be used to transform virtually any crop plant cell under suitable Transformed cells can be regenerated into whole plants such that the conditions. nucleotide sequence of the invention confer insect resistance to the transgenic plants.

Example 22: Modification of Coding Sequences and Adjacent Sequences

The nucleotide sequences described in this application can be modified for expression in transgenic plant hosts. A host plant expressing the nucleotide sequences and

WO 99/42589 PCT/EP99/01015

- 45 -

which produces the insecticidal toxins in its cells has enhanced resistance to insect attack and is thus better equipped to withstand crop losses associated with such attack.

The transgenic expression in plants of genes derived from microbial sources may require the modification of those genes to achieve and optimize their expression in plants. In particular, bacterial ORFs which encode separate enzymes but which are encoded by the same transcript in the native microbe are best expressed in plants on separate transcripts. To achieve this, each microbial ORF is isolated individually and cloned within a cassette which provides a plant promoter sequence at the 5' end of the ORF and a plant transcriptional terminator at the 3' end of the ORF. The isolated ORF sequence preferably includes the initiating ATG codon and the terminating STOP codon but may include additional sequence beyond the initiating ATG and the STOP codon. In addition, the ORF may be truncated, but still retain the required activity; for particularly long ORFs, truncated versions which retain activity may be preferable for expression in transgenic organisms. By "plant promoter" and "plant transcriptional terminator" it is intended to mean promoters and transcriptional terminators which operate within plant cells. This includes promoters and transcription terminators which may be derived from non-plant sources such as viruses (an example is the Cauliflower Mosaic Virus).

In some cases, modification to the ORF coding sequences and adjacent sequence is not required. It is sufficient to isolate a fragment containing the ORF of interest and to insert it downstream of a plant promoter. For example, Gaffney et al. (Science 261: 754-756 (1993)) have expressed the *Pseudomonas nahG* gene in transgenic plants under the control of the CaMV 35S promoter and the CaMV tml terminator successfully without modification of the coding sequence and with x bp of the *Pseudomonas* gene upstream of the ATG still attached, and y bp downstream of the STOP codon still attached to the *nahG* ORF. Preferably as little adjacent microbial sequence should be left attached upstream of the ATG and downstream of the STOP codon. In practice, such construction may depend on the availability of restriction sites.

In other cases, the expression of genes derived from microbial sources may provide problems in expression. These problems have been well characterized in the art and are particularly common with genes derived from certain sources such as *Bacillus*. These problems may apply to the nucleotide sequence of this invention and the modification of these genes can be undertaken using techniques now well known in the art. The following problems may be encountered:

WO 99/42589 PCT/EP99/01015

- 46 -

1. Codon Usage.

The preferred codon usage in plants differs from the preferred codon usage in certain microorganisms. Comparison of the usage of codons within a cloned microbial ORF to usage in plant genes (and in particular genes from the target plant) will enable an identification of the codons within the ORF which should preferably be changed. Typically plant evolution has tended towards a strong preference of the nucleotides C and G in the third base position of monocotyledons, whereas dicotyledons often use the nucleotides A or T at this position. By modifying a gene to incorporate preferred codon usage for a particular target transgenic species, many of the problems described below for GC/AT content and illegitimate splicing will be overcome.

2. GC/AT Content.

Plant genes typically have a GC content of more than 35%. ORF sequences which are rich in A and T nucleotides can cause several problems in plants. Firstly, motifs of ATTTA are believed to cause destabilization of messages and are found at the 3' end of many short-lived mRNAs. Secondly, the occurrence of polyadenylation signals such as AATAAA at inappropriate positions within the message is believed to cause premature truncation of transcription. In addition, monocotyledons may recognize AT-rich sequences as splice sites (see below).

3. Sequences Adjacent to the Initiating Methionine.

Plants differ from microorganisms in that their messages do not possess a defined ribosome binding site. Rather, it is believed that ribosomes attach to the 5' end of the message and scan for the first available ATG at which to start translation. Nevertheless, it is believed that there is a preference for certain nucleotides adjacent to the ATG and that expression of microbial genes can be enhanced by the inclusion of a eukaryotic consensus translation initiator at the ATG. Clontech (1993/1994 catalog, page 210, incorporated herein by reference) have suggested one sequence as a consensus translation initiator for the expression of the *E. coli uidA* gene in plants. Further, Joshi (NAR 15: 6643-6653 (1987), incorporated herein by reference) has compared many plant sequences adjacent to the ATG and suggests another consensus sequence. In situations where difficulties are encountered in the expression of microbial ORFs in plants, inclusion of one of these sequences at the initiating ATG may improve translation. In such cases the last three

nucleotides of the consensus may not be appropriate for inclusion in the modified sequence due to their modification of the second AA residue. Preferred sequences adjacent to the initiating methionine may differ between different plant species. A survey of 14 maize genes located in the GenBank database provided the following results:

Position Be	fore the	Initiating	ATG in	14	Maize	Genes:

	<u>-10</u>	<u>-9</u>	<u>-8</u>	<u>-7</u>	<u>-6</u>	<u>-5</u>	<u>-4</u>	<u>-3</u>	<u>-2</u>	<u>-1</u>
С	3	8	4	6	2	5	6	0	10	7
T	3	0	3	4	3	2	1	1	1	0
A	2	3	1	4	3	2	3	7	2	3
G	6	3	6	0	6	5	4	6	1	5

This analysis can be done for the desired plant species into which the nucleotide sequence is being incorporated, and the sequence adjacent to the ATG modified to incorporate the preferred nucleotides.

4. Removal of Illegitimate Splice Sites.

Genes cloned from non-plant sources and not optimized for expression in plants may also contain motifs which may be recognized in plants as 5' or 3' splice sites, and be cleaved, thus generating truncated or deleted messages. These sites can be removed using the techniques well known in the art.

Techniques for the modification of coding sequences and adjacent sequences are well known in the art. In cases where the initial expression of a microbial ORF is low and it is deemed appropriate to make alterations to the sequence as described above, then the construction of synthetic genes can be accomplished according to methods well known in the art. These are, for example, described in the published patent disclosures EP 0 385 962 (to Monsanto), EP 0 359 472 (to Lubrizol) and WO 93/07278 (to Ciba-Geigy), all of which are incorporated herein by reference. In most cases it is preferable to assay the expression of gene constructions using transient assay protocols (which are well known in the art) prior to their transfer to transgenic plants.

Example 23: Construction of Plant Expr ssion Cass tt s

Coding sequences intended for expression in transgenic plants are first assembled in expression cassettes behind a suitable promoter expressible in plants. The expression cassettes may also comprise any further sequences required or selected for the expression of the transgene. Such sequences include, but are not restricted to, transcription terminators, extraneous sequences to enhance expression such as introns, vital sequences, and sequences intended for the targeting of the gene product to specific organelles and cell compartments. These expression cassettes can then be easily transferred to the plant transformation vectors described below. The following is a description of various components of typical expression cassettes.

1. Promoters

The selection of the promoter used in expression cassettes will determine the spatial and temporal expression pattern of the transgene in the transgenic plant. Selected promoters will express transgenes in specific cell types (such as leaf epidermal cells, mesophyll cells, root cortex cells) or in specific tissues or organs (roots, leaves or flowers, for example) and the selection will reflect the desired location of accumulation of the gene product. Alternatively, the selected promoter may drive expression of the gene under various inducing conditions. Promoters vary in their strength, i.e., ability to promote transcription. Depending upon the host cell system utilized, any one of a number of suitable promoters can be used, including the gene's native promoter. The following are non-limiting examples of promoters that may be used in expression cassettes.

a. Constitutive Expression, the Ubiquitin Promoter:

Ubiquitin is a gene product known to accumulate in many cell types and its promoter has been cloned from several species for use in transgenic plants (e.g. sunflower - Binet et al. Plant Science 79: 87-94 (1991); maize - Christensen et al. Plant Molec. Biol. 12: 619-632 (1989); and Arabidopsis - Norris et al., Plant Mol. Biol. 21:895-906 (1993)). The maize ubiquitin promoter has been developed in transgenic monocot systems and its sequence and vectors constructed for monocot transformation are disclosed in the patent publication EP 0 342 926 (to Lubrizol) which is herein incorporated by reference. Taylor et al. (Plant Cell Rep. 12: 491-495 (1993)) describe a vector (pAHC25) that comprises the maize ubiquitin promoter and first intron and its high activity in cell suspensions of numerous

monocotyledons when introduced via microprojectile bombardment. The *Arabidopsis* ubiquitin promoter is ideal for use with the nucleotide sequences of the present invention. The ubiquitin promoter is suitable for gene expression in transgenic plants, both monocotyledons and dicotyledons. Suitable vectors are derivatives of pAHC25 or any of the transformation vectors described in this application, modified by the introduction of the appropriate ubiquitin promoter and/or intron sequences.

- 49 -

b. Constitutive Expression, the CaMV 35S Promoter:

Construction of the plasmid pCGN1761 is described in the published patent application EP 0 392 225 (Example 23), which is hereby incorporated by reference. pCGN1761 contains the "double" CaMV 35S promoter and the tml transcriptional terminator with a unique EcoRI site between the promoter and the terminator and has a pUC-type backbone. A derivative of pCGN1761 is constructed which has a modified polylinker which includes NotI and XhoI sites in addition to the existing EcoRI site. This derivative is designated pCGN1761ENX. pCGN1761ENX is useful for the cloning of cDNA sequences or coding sequences (including microbial ORF sequences) within its polylinker for the purpose of their expression under the control of the 35S promoter in transgenic plants. The entire 35S promoter-coding sequence-tml terminator cassette of such a construction can be excised by HindIII, Sphl, Sall, and Xbal sites 5' to the promoter and Xbal, BamHI and Ball sites 3' to the terminator for transfer to transformation vectors such as those described below. Furthermore, the double 35S promoter fragment can be removed by 5' excision with HindIII, SphI, Sall, Xbal, or PstI, and 3' excision with any of the polylinker restriction sites (EcoRI, NotI or XhoI) for replacement with another promoter. If desired, modifications around the cloning sites can be made by the introduction of sequences that may enhance translation. This is particularly useful when overexpression is desired. For example, pCGN1761ENX may be modified by optimization of the translational initiation site as described in Example 37 of U.S. Patent No. 5,639,949, incorporated herein by reference.

c. Constitutive Expression, the Actin Promoter:

Several isoforms of actin are known to be expressed in most cell types and consequently the actin promoter is a good choice for a constitutive promoter. In particular, the promoter from the rice *Actl* gene has been cloned and characterized (McElroy *et al.* Plant Cell 2: 163-171 (1990)). A 1.3kb fragment of the promoter was found to contain all

the regulatory elements required for expression in rice protoplasts. Furthermore, numerous expression vectors based on the ActI promoter have been constructed specifically for use in monocotyledons (McElroy et al. Mol. Gen. Genet. 231: 150-160 (1991)). These incorporate the ActI-intron 1, AdhI 5' flanking sequence and AdhI-intron 1 (from the maize alcohol dehydrogenase gene) and sequence from the CaMV 35S promoter. Vectors showing highest expression were fusions of 35S and Actl intron or the Actl 5' flanking sequence and the ActI intron. Optimization of sequences around the initiating ATG (of the GUS reporter gene) also enhanced expression. The promoter expression cassettes described by McElroy et al. (Mol. Gen. Genet. 231: 150-160 (1991)) can be easily modified for gene expression and are particularly suitable for use in monocotyledonous hosts. For example, promotercontaining fragments is removed from the McElroy constructions and used to replace the double 35S promoter in pCGN1761ENX, which is then available for the insertion of specific gene sequences. The fusion genes thus constructed can then be transferred to appropriate transformation vectors. In a separate report, the rice Actl promoter with its first intron has also been found to direct high expression in cultured barley cells (Chibbar et al. Plant Cell Rep. <u>12</u>: 506-509 (1993)).

d. Inducible Expression, the PR-1 Promoter:

The double 35S promoter in pCGN1761ENX may be replaced with any other promoter of choice that will result in suitably high expression levels. By way of example, one of the chemically regulatable promoters described in U.S. Patent No. 5,614,395 may replace the double 35S promoter. The promoter of choice is preferably excised from its source by restriction enzymes, but can alternatively be PCR-amplified using primers that carry appropriate terminal restriction sites. Should PCR-amplification be undertaken, then the promoter should be re-sequenced to check for amplification errors after the cloning of the amplified promoter in the target vector. The chemically/pathogen regulatable tobacco PR-1a promoter is cleaved from plasmid pCIB1004 (for construction, see example 21 of EP 0 332 104, which is hereby incorporated by reference) and transferred to plasmid pCGN1761ENX (Uknes et al., 1992). pCIB1004 is cleaved with *Ncol* and the resultant 3' overhang of the linearized fragment is rendered blunt by treatment with T4 DNA polymerase. The fragment is then cleaved with *HindIII* and the resultant PR-1a promotercontaining fragment is gel purified and cloned into pCGN1761ENX from which the double 35S promoter has been removed. This is done by cleavage with *Xhol* and blunting with T4

- 51 -

PCT/EP99/01015

polymerase, followed by cleavage with *HindIII* and isolation of the larger vector-terminator containing fragment into which the pCIB1004 promoter fragment is cloned. This generates a pCGN1761ENX derivative with the PR-1a promoter and the *tml* terminator and an intervening polylinker with unique *EcoRI* and *NotI* sites. The selected coding sequence can be inserted into this vector, and the fusion products (*i.e.* promoter-gene-terminator) can subsequently be transferred to any selected transformation vector, including those described *infra*. Various chemical regulators may be employed to induce expression of the selected coding sequence in the plants transformed according to the present invention, including the benzothiadiazole, isonicotinic acid, and salicylic acid compounds disclosed in U.S. Patent Nos. 5,523,311 and 5,614,395.

e. Inducible Expression, an Ethanol-Inducible Promoter:

A promoter inducible by certain alcohols or ketones, such as ethanol, may also be used to confer inducible expression of a coding sequence of the present invention. Such a promoter is for example the *alcA* gene promoter from *Aspergillus nidulans* (Caddick et al. (1998) *Nat. Biotechnol* 16:177-180). In *A. nidulans*, the *alcA* gene encodes alcohol dehydrogenase I, the expression of which is regulated by the AlcR transcription factors in presence of the chemical inducer. For the purposes of the present invention, the CAT coding sequences in plasmid palcA:CAT comprising a *alcA* gene promoter sequence fused to a minimal 35S promoter (Caddick et al. (1998) *Nat. Biotechnol* 16:177-180) are replaced by a coding sequence of the present invention to form an expression cassette having the coding sequence under the control of the *alcA* gene promoter. This is carried out using methods well known in the art.

f. Inducible Expression, a Glucocorticoid-Inducible Promoter:

Induction of expression of a nucleic acid sequence of the present invention using systems based on steroid hormones is also contemplated. For example, a glucocorticoid-mediated induction system is used (Aoyama and Chua (1997) *The Plant Journal* 11: 605-612) and gene expression is induced by application of a glucocorticoid, for example a synthetic glucocorticoid, preferably dexamethasone, preferably at a concentration ranging from 0.1mM to 1mM, more preferably from 10mM to 100mM. For the purposes of the present invention, the luciferase gene sequences are replaced by a nucleic acid sequence of the invention to form an expression cassette having a nucleic acid sequence of the

WO 99/42589 PCT/EP99/01015

- 52 -

invention under the control of six copies of the GAL4 upstream activating sequences fused to the 35S minimal promoter. This is carried out using methods well known in the art. The trans-acting factor comprises the GAL4 DNA-binding domain (Keegan et al. (1986) *Science* 231: 699-704) fused to the transactivating domain of the herpes viral protein VP16 (Triezenberg et al. (1988) *Genes Devel.* 2: 718-729) fused to the hormone-binding domain of the rat glucocorticoid receptor (Picard et al. (1988) *Cell* 54: 1073-1080). The expression of the fusion protein is controlled by any promoter suitable for expression in plants known in the art or described here. This expression cassette is also comprised in the plant comprising a nucleic acid sequence of the invention fused to the 6xGAL4/minimal promoter. Thus, tissue- or organ-specificity of the fusion protein is achieved leading to inducible tissue- or organ-specificity of the insecticidal toxin.

g. Root Specific Expression:

Another pattern of gene expression is root expression. A suitable root promoter is described by de Framond (FEBS 290: 103-106 (1991)) and also in the published patent application EP 0 452 269, which is herein incorporated by reference. This promoter is transferred to a suitable vector such as pCGN1761ENX for the insertion of a selected gene and subsequent transfer of the entire promoter-gene-terminator cassette to a transformation vector of interest.

h. Wound-Inducible Promoters:

Wound-inducible promoters may also be suitable for gene expression. Numerous such promoters have been described (*e.g.* Xu *et al.* Plant Molec. Biol. <u>22</u>: 573-588 (1993), Logemann *et al.* Plant Cell <u>1</u>: 151-158 (1989), Rohrmeier & Lehle, Plant Molec. Biol. <u>22</u>: 783-792 (1993), Firek *et al.* Plant Molec. Biol. <u>22</u>: 129-142 (1993), Warner *et al.* Plant J. <u>3</u>: 191-201 (1993)) and all are suitable for use with the instant invention. Logemann *et al.* describe the 5' upstream sequences of the dicotyledonous potato *wunl* gene. Xu *et al.* show that a wound-inducible promoter from the dicotyledon potato (*pin2*) is active in the monocotyledon rice. Further, Rohrmeier & Lehle describe the cloning of the maize *Wipl* cDNA which is wound induced and which can be used to isolate the cognate promoter using standard techniques. Similar, Firek *et al.* and Warner *et al.* have described a wound-induced gene from the monocotyledon *Asparagus officinalis*, which is expressed at local wound and pathogen invasion sites. Using cloning techniques well known in the art, these

promoters can be transferred to suitable vectors, fused to the genes pertaining to this invention, and used to express these genes at the sites of plant wounding.

i. Pith-Preferred Expression:

Patent Application WO 93/07278, which is herein incorporated by reference, describes the isolation of the maize *trpA* gene, which is preferentially expressed in pith cells. The gene sequence and promoter extending up to -1726 bp from the start of transcription are presented. Using standard molecular biological techniques, this promoter, or parts thereof, can be transferred to a vector such as pCGN1761 where it can replace the 35S promoter and be used to drive the expression of a foreign gene in a pith-preferred manner. In fact, fragments containing the pith-preferred promoter or parts thereof can be transferred to any vector and modified for utility in transgenic plants.

j. Leaf-Specific Expression:

A maize gene encoding phosphoenol carboxylase (PEPC) has been described by Hudspeth & Grula (Plant Molec Biol 12: 579-589 (1989)). Using standard molecular biological techniques the promoter for this gene can be used to drive the expression of any gene in a leaf-specific manner in transgenic plants.

k. Pollen-Specific Expression:

WO 93/07278 describes the isolation of the maize calcium-dependent protein kinase (CDPK) gene which is expressed in pollen cells. The gene sequence and promoter extend up to 1400 bp from the start of transcription. Using standard molecular biological techniques, this promoter or parts thereof, can be transferred to a vector such as pCGN1761 where it can replace the 35S promoter and be used to drive the expression of a nucleic acid sequence of the invention in a pollen-specific manner.

2. Transcriptional Terminators

A variety of transcriptional terminators are available for use in expression cassettes. These are responsible for the termination of transcription beyond the transgene and its correct polyadenylation. Appropriate transcriptional terminators are those that are known to function in plants and include the CaMV 35S terminator, the *tml* terminator, the nopaline synthase terminator and the pea *rbcS* E9 terminator. These can be used in both

monocotyledons and dicotyledons. In addition, a gene's native transcription terminator may be used.

3. Sequences for the Enhancement or Regulation of Expression

Numerous sequences have been found to enhance gene expression from within the transcriptional unit and these sequences can be used in conjunction with the genes of this invention to increase their expression in transgenic plants.

Various intron sequences have been shown to enhance expression, particularly in monocotyledonous cells. For example, the introns of the maize *AdhI* gene have been found to significantly enhance the expression of the wild-type gene under its cognate promoter when introduced into maize cells. Intron 1 was found to be particularly effective and enhanced expression in fusion constructs with the chloramphenicol acetyltransferase gene (Callis *et al.*, Genes Develop. 1: 1183-1200 (1987)). In the same experimental system, the intron from the maize *bronze1* gene had a similar effect in enhancing expression. Intron sequences have been routinely incorporated into plant transformation vectors, typically within the non-translated leader.

A number of non-translated leader sequences derived from viruses are also known to enhance expression, and these are particularly effective in dicotyledonous cells. Specifically, leader sequences from Tobacco Mosaic Virus (TMV, the "W-sequence"), Maize Chlorotic Mottle Virus (MCMV), and Alfalfa Mosaic Virus (AMV) have been shown to be effective in enhancing expression (e.g. Gallie et al. Nucl. Acids Res. 15: 8693-8711 (1987); Skuzeski et al. Plant Molec. Biol. 15: 65-79 (1990)).

4. Targeting of the Gene Product Within the Cell

Various mechanisms for targeting gene products are known to exist in plants and the sequences controlling the functioning of these mechanisms have been characterized in some detail. For example, the targeting of gene products to the chloroplast is controlled by a signal sequence found at the amino terminal end of various proteins which is cleaved during chloroplast import to yield the mature protein (*e.g.* Comai *et al.* J. Biol. Chem. <u>263</u>: 15104-15109 (1988)). These signal sequences can be fused to heterologous gene products to effect the import of heterologous products into the chloroplast (van den Broeck, et al. Nature <u>313</u>: 358-363 (1985)). DNA encoding for appropriate signal sequences can be isolated from the 5' end of the cDNAs encoding the RUBISCO protein, the CAB protein, the

EPSP synthase enzyme, the GS2 protein and many other proteins which are known to be chloroplast localized. *See also*, the section entitled "Expression With Chloroplast Targeting" in Example 37 of U.S. Patent No. 5,639,949.

Other gene products are localized to other organelles such as the mitochondrion and the peroxisome (e.g. Unger et al. Plant Molec. Biol. 13: 411-418 (1989)). The cDNAs encoding these products can also be manipulated to effect the targeting of heterologous gene products to these organelles. Examples of such sequences are the nuclear-encoded ATPases and specific aspartate amino transferase isoforms for mitochondria. Targeting cellular protein bodies has been described by Rogers et al. (Proc. Natl. Acad. Sci. USA 82: 6512-6516 (1985)).

In addition, sequences have been characterized which cause the targeting of gene products to other cell compartments. Amino terminal sequences are responsible for targeting to the ER, the apoplast, and extracellular secretion from aleurone cells (Koehler & Ho, Plant Cell 2: 769-783 (1990)). Additionally, amino terminal sequences in conjunction with carboxy terminal sequences are responsible for vacuolar targeting of gene products (Shinshi *et al.* Plant Molec. Biol. 14: 357-368 (1990)).

By the fusion of the appropriate targeting sequences described above to transgene sequences of interest it is possible to direct the transgene product to any organelle or cell compartment. For chloroplast targeting, for example, the chloroplast signal sequence from the RUBISCO gene, the CAB gene, the EPSP synthase gene, or the GS2 gene is fused in frame to the amino terminal ATG of the transgene. The signal sequence selected should include the known cleavage site, and the fusion constructed should take into account any amino acids after the cleavage site which are required for cleavage. In some cases this requirement may be fulfilled by the addition of a small number of amino acids between the cleavage site and the transgene ATG or, alternatively, replacement of some amino acids within the transgene sequence. Fusions constructed for chloroplast import can be tested for efficacy of chloroplast uptake by in vitro translation of in vitro transcribed constructions followed by in vitro chloroplast uptake using techniques described by Bartlett et al. In: Edelmann et al. (Eds.) Methods in Chloroplast Molecular Biology, Elsevier pp 1081-1091 (1982) and Wasmann et al. Mol. Gen. Genet. 205: 446-453 (1986). These construction techniques are well known in the art and are equally applicable to mitochondria and peroxisomes.

The above-described mechanisms for cellular targeting can be utilized not only in conjunction with their cognate promoters, but also in conjunction with heterologous promoters so as to effect a specific cell-targeting goal under the transcriptional regulation of a promoter that has an expression pattern different to that of the promoter from which the targeting signal derives.

Example 24: Construction of Plant Transformation Vectors

Numerous transformation vectors available for plant transformation are known to those of ordinary skill in the plant transformation arts, and the genes pertinent to this invention can be used in conjunction with any such vectors. The selection of vector will depend upon the preferred transformation technique and the target species for transformation. For certain target species, different antibiotic or herbicide selection markers may be preferred. Selection markers used routinely in transformation include the *nptll* gene, which confers resistance to kanamycin and related antibiotics (Messing & Vierra. Gene 19: 259-268 (1982); Bevan et al., Nature 304:184-187 (1983)), the *bar* gene, which confers resistance to the herbicide phosphinothricin (White et al., Nucl. Acids Res 18: 1062 (1990), Spencer et al. Theor. Appl. Genet 79: 625-631 (1990)), the *hph* gene, which confers resistance to the antibiotic hygromycin (Blochinger & Diggelmann, Mol Cell Biol 4: 2929-2931), and the *dhfr* gene, which confers resistance to methatrexate (Bourouis et al., EMBO J. 2(7): 1099-1104 (1983)), and the EPSPS gene, which confers resistance to glyphosate (U.S. Patent Nos. 4,940,935 and 5,188,642).

1. Vectors Suitable for Agrobacterium Transformation

Many vectors are available for transformation using *Agrobacterium tumefaciens*. These typically carry at least one T-DNA border sequence and include vectors such as pBIN19 (Bevan, Nucl. Acids Res. (1984)) and pXYZ. Below, the construction of two typical vectors suitable for *Agrobacterium* transformation is described.

a. pClB200 and pClB2001:

The binary vectors pclB200 and pClB2001 are used for the construction of recombinant vectors for use with *Agrobacterium* and are constructed in the following manner. pTJS75kan is created by *Narl* digestion of pTJS75 (Schmidhauser & Helinski, J.

Bacteriol. 164: 446-455 (1985)) allowing excision of the tetracycline-resistance gene. followed by insertion of an Accl fragment from pUC4K carrying an NPTII (Messing & Vierra. Gene 19: 259-268 (1982): Bevan et al., Nature 304: 184-187 (1983): McBride et al., Plant Molecular Biology 14: 266-276 (1990)). Xhol linkers are ligated to the EcoRV fragment of PCIB7 which contains the left and right T-DNA borders, a plant selectable nos/nptll chimeric gene and the pUC polylinker (Rothstein et al., Gene 53: 153-161 (1987)), and the Xholdigested fragment are cloned into Sall-digested pTJS75kan to create pClB200 (see also EP 0 332 104, example 19). pCIB200 contains the following unique polylinker restriction sites: EcoRI, Sstl, KpnI, BgIII, XbaI, and Sall. pCIB2001 is a derivative of pCIB200 created by the insertion into the polylinker of additional restriction sites. Unique restriction sites in the polylinker of pClB2001 are EcoRI, Sstl, KpnI, BgIII, XbaI, SalI, MluI, BclI, AvrII, ApaI, HpaI, and Stul. pCIB2001, in addition to containing these unique restriction sites also has plant and bacterial kanamycin selection, left and right T-DNA borders for Agrobacterium-mediated transformation, the RK2-derived trfA function for mobilization between E. coli and other hosts, and the OriT and OriV functions also from RK2. The pCIB2001 polylinker is suitable for the cloning of plant expression cassettes containing their own regulatory signals.

b. pCIB10 and Hygromycin Selection Derivatives thereof:

The binary vector pCIB10 contains a gene encoding kanamycin resistance for selection in plants and T-DNA right and left border sequences and incorporates sequences from the wide host-range plasmid pRK252 allowing it to replicate in both *E. coli* and *Agrobacterium*. Its construction is described by Rothstein *et al.* (Gene <u>53</u>: 153-161 (1987)). Various derivatives of pCIB10 are constructed which incorporate the gene for hygromycin B phosphotransferase described by Gritz *et al.* (Gene <u>25</u>: 179-188 (1983)). These derivatives enable selection of transgenic plant cells on hygromycin only (pCIB743), or hygromycin and kanamycin (pCIB715, pCIB717).

2. Vectors Suitable for non-Agrobacterium Transformation

Transformation without the use of *Agrobacterium tumefaciens* circumvents the requirement for T-DNA sequences in the chosen transformation vector and consequently vectors lacking these sequences can be utilized in addition to vectors such as the ones described above which contain T-DNA sequences. Transformation techniques that do not rely on *Agrobacterium* include transformation via particle bombardment, protoplast uptake

(e.g. PEG and electroporation) and microinjection. The choice of vector depends largely on the preferred selection for the species being transformed. Below, the construction of typical vectors suitable for non-Agrobacterium transformation is described.

a. pCIB3064:

pClB3064 is a pUC-derived vector suitable for direct gene transfer techniques in combination with selection by the herbicide basta (or phosphinothricin). The plasmid pCIB246 comprises the CaMV 35S promoter in operational fusion to the E. coli GUS gene and the CaMV 35S transcriptional terminator and is described in the PCT published application WO 93/07278. The 35S promoter of this vector contains two ATG sequences 5' of the start site. These sites are mutated using standard PCR techniques in such a way as to remove the ATGs and generate the restriction sites Sspl and Pvull. The new restriction sites are 96 and 37 bp away from the unique Sall site and 101 and 42 bp away from the actual start site. The resultant derivative of pCIB246 is designated pCIB3025. The GUS gene is then excised from pClB3025 by digestion with Sall and Sacl, the termini rendered blunt and religated to generate plasmid pCIB3060. The plasmid pJIT82 is obtained from the John Innes Centre, Norwich and the a 400 bp Smal fragment containing the bar gene from Streptomyces viridochromogenes is excised and inserted into the Hpal site of pCIB3060 (Thompson et al. EMBO J 6: 2519-2523 (1987)). This generated pCIB3064, which comprises the bar gene under the control of the CaMV 35S promoter and terminator for herbicide selection, a gene for ampicillin resistance (for selection in E. coli) and a polylinker with the unique sites Sphl, Pstl, Hindlll, and BamHl. This vector is suitable for the cloning of plant expression cassettes containing their own regulatory signals.

b. pSOG19 and pSOG35:

pSOG35 is a transformation vector that utilizes the *E. coli* gene dihydrofolate reductase (DFR) as a selectable marker conferring resistance to methotrexate. PCR is used to amplify the 35S promoter (-800 bp), intron 6 from the maize Adh1 gene (-550 bp) and 18 bp of the GUS untranslated leader sequence from pSOG10. A 250-bp fragment encoding the *E. coli* dihydrofolate reductase type II gene is also amplified by PCR and these two PCR fragments are assembled with a *SacI-PstI* fragment from pB1221 (Clontech) which comprises the pUC19 vector backbone and the nopaline synthase terminator. Assembly of these fragments generates pSOG19 which contains the 35S promoter in fusion

with the intron 6 sequence, the GUS leader, the DHFR gene and the nopaline synthase terminator. Replacement of the GUS leader in pSOG19 with the leader sequence from Maize Chlorotic Mottle Virus (MCMV) generates the vector pSOG35. pSOG19 and pSOG35 carry the pUC gene for ampicillin resistance and have *HindIII*, *SphI*, *PstI* and *EcoRI* sites available for the cloning of foreign substances.

Example 25: Transformation

Once a nucleic acid sequence of the invention has been cloned into an expression system, it is transformed into a plant cell. Methods for transformation and regeneration of plants are well known in the art. For example, Ti plasmid vectors have been utilized for the delivery of foreign DNA, as well as direct DNA uptake, liposomes, electroporation, microinjection, and microprojectiles. In addition, bacteria from the genus *Agrobacterium* can be utilized to transform plant cells. Below are descriptions of representative techniques for transforming both dicotyledonous and monocotyledonous plants.

1. Transformation of Dicotyledons

Transformation techniques for dicotyledons are well known in the art and include Agrobacterium-based techniques and techniques that do not require Agrobacterium. Non-Agrobacterium techniques involve the uptake of exogenous genetic material directly by protoplasts or cells. This can be accomplished by PEG or electroporation mediated uptake, particle bombardment-mediated delivery, or microinjection. Examples of these techniques are described by Paszkowski et al., EMBO J 3: 2717-2722 (1984), Potrykus et al., Mol. Gen. Genet. 199: 169-177 (1985), Reich et al., Biotechnology 4: 1001-1004 (1986), and Klein et al., Nature 327: 70-73 (1987). In each case the transformed cells are regenerated to whole plants using standard techniques known in the art.

Agrobacterium-mediated transformation is a preferred technique for transformation of dicotyledons because of its high efficiency of transformation and its broad utility with many different species. Agrobacterium transformation typically involves the transfer of the binary vector carrying the foreign DNA of interest (e.g. pCIB200 or pCIB2001) to an appropriate Agrobacterium strain which may depend of the complement of vir genes carried by the host Agrobacterium strain either on a co-resident Ti plasmid or chromosomally (e.g. strain CIB542 for pCIB200 and pCIB2001 (Uknes et al. Plant Cell 5: 159-169 (1993)). The

transfer of the recombinant binary vector to *Agrobacterium* is accomplished by a triparental mating procedure using *E. coli* carrying the recombinant binary vector, a helper *E. coli* strain which carries a plasmid such as pRK2013 and which is able to mobilize the recombinant binary vector to the target *Agrobacterium* strain. Alternatively, the recombinant binary vector can be transferred to *Agrobacterium* by DNA transformation (Höfgen & Willmitzer, Nucl. Acids Res. 16: 9877 (1988)).

Transformation of the target plant species by recombinant *Agrobacterium* usually involves co-cultivation of the *Agrobacterium* with explants from the plant and follows protocols well known in the art. Transformed tissue is regenerated on selectable medium carrying the antibiotic or herbicide resistance marker present between the binary plasmid T-DNA borders.

Another approach to transforming plant cells with a gene involves propelling inert or biologically active particles at plant tissues and cells. This technique is disclosed in U.S. Patent Nos. 4,945,050, 5,036,006, and 5,100,792 all to Sanford et al. Generally, this procedure involves propelling inert or biologically active particles at the cells under conditions effective to penetrate the outer surface of the cell and afford incorporation within the interior thereof. When inert particles are utilized, the vector can be introduced into the cell by coating the particles with the vector containing the desired gene. Alternatively, the target cell can be surrounded by the vector so that the vector is carried into the cell by the wake of the particle. Biologically active particles (e.g., dried yeast cells, dried bacterium or a bacteriophage, each containing DNA sought to be introduced) can also be propelled into plant cell tissue.

2. Transformation of Monocotyledons

Transformation of most monocotyledon species has now also become routine. Preferred techniques include direct gene transfer into protoplasts using PEG or electroporation techniques, and particle bombardment into callus tissue. Transformations can be undertaken with a single DNA species or multiple DNA species (*i.e.* cotransformation) and both these techniques are suitable for use with this invention. Cotransformation may have the advantage of avoiding complete vector construction and of generating transgenic plants with unlinked loci for the gene of interest and the selectable marker, enabling the removal of the selectable marker in subsequent generations, should this be regarded desirable. However, a disadvantage of the use of co-transformation is the

less than 100% frequency with which separate DNA species are integrated into the genome (Schocher *et al.* Biotechnology <u>4</u>: 1093-1096 (1986)).

Patent Applications EP 0 292 435, EP 0 392 225, and WO 93/07278 describe techniques for the preparation of callus and protoplasts from an elite inbred line of maize, transformation of protoplasts using PEG or electroporation, and the regeneration of maize plants from transformed protoplasts. Gordon-Kamm *et al.* (Plant Cell <u>2</u>: 603-618 (1990)) and Fromm *et al.* (Biotechnology <u>8</u>: 833-839 (1990)) have published techniques for transformation of A188-derived maize line using particle bombardment. Furthermore, WO 93/07278 and Koziel *et al.* (Biotechnology <u>11</u>: 194-200 (1993)) describe techniques for the transformation of elite inbred lines of maize by particle bombardment. This technique utilizes immature maize embryos of 1.5-2.5 mm length excised from a maize ear 14-15 days after pollination and a PDS-1000He Biolistics device for bombardment.

Transformation of rice can also be undertaken by direct gene transfer techniques utilizing protoplasts or particle bombardment. Protoplast-mediated transformation has been described for *Japonica*-types and *Indica*-types (Zhang *et al.* Plant Cell Rep 7: 379-384 (1988); Shimamoto *et al.* Nature 338: 274-277 (1989); Datta *et al.* Biotechnology 8: 736-740 (1990)). Both types are also routinely transformable using particle bombardment (Christou *et al.* Biotechnology 9: 957-962 (1991)). Furthermore, WO 93/21335 describes techniques for the transformation of rice via electroporation.

Patent Application EP 0 332 581 describes techniques for the generation, transformation and regeneration of Pooideae protoplasts. These techniques allow the transformation of *Dactylis* and wheat. Furthermore, wheat transformation has been described by Vasil *et al.* (Biotechnology 10: 667-674 (1992)) using particle bombardment into cells of type C long-term regenerable callus, and also by Vasil *et al.* (Biotechnology 11: 1553-1558 (1993)) and Weeks *et al.* (Plant Physiol. 102: 1077-1084 (1993)) using particle bombardment of immature embryos and immature embryo-derived callus. A preferred technique for wheat transformation, however, involves the transformation of wheat by particle bombardment of immature embryos and includes either a high sucrose or a high maltose step prior to gene delivery. Prior to bombardment, any number of embryos (0.75-1 mm in length) are plated onto MS medium with 3% sucrose (Murashiga & Skoog, Physiologia Plantarum 15: 473-497 (1962)) and 3 mg/l 2,4-D for induction of somatic embryos, which is allowed to proceed in the dark. On the chosen day of bombardment, embryos are removed from the induction medium and placed onto the osmoticum (*i.e.*

induction medium with sucrose or maltose added at the desired concentration, typically 15%). The embryos are allowed to plasmolyze for 2-3 h and are then bombarded. Twenty embryos per target plate is typical, although not critical. An appropriate gene-carrying plasmid (such as pClB3064 or pSG35) is precipitated onto micrometer size gold particles using standard procedures. Each plate of embryos is shot with the DuPont Biolistics® helium device using a burst pressure of ~1000 psi using a standard 80 mesh screen. After bombardment, the embryos are placed back into the dark to recover for about 24 h (still on osmoticum). After 24 hrs, the embryos are removed from the osmoticum and placed back onto induction medium where they stay for about a month before regeneration. Approximately one month later the embryo explants with developing embryogenic callus are transferred to regeneration medium (MS + 1 mg/liter NAA, 5 mg/liter GA), further containing the appropriate selection agent (10 mg/l basta in the case of pClB3064 and 2 mg/l methotrexate in the case of pSOG35). After approximately one month, developed shoots are transferred to larger sterile containers known as "GA7s" which contain half-strength MS, 2% sucrose, and the same concentration of selection agent.

Tranformation of monocotyledons using *Agrobacterium* has also been described. *See,* WO 94/00977 and U.S. Patent No. 5,591,616, both of which are incorporated herein by reference.

E. Breeding and Seed Production

Example 26: Breeding

The plants obtained via tranformation with a nucleic acid sequence of the present invention can be any of a wide variety of plant species, including those of monocots and dicots; however, the plants used in the method of the invention are preferably selected from the list of agronomically important target crops set forth *supra*. The expression of a gene of the present invention in combination with other characteristics important for production and quality can be incorporated into plant lines through breeding. Breeding approaches and techniques are known in the art. See, for example, Welsh J. R., *Fundamentals of Plant Genetics and Breeding*, John Wiley & Sons, NY (1981); *Crop Breeding*, Wood D. R. (Ed.) American Society of Agronomy Madison, Wisconsin (1983); Mayo O., *The Theory of Plant Breeding*, Second Edition, Clarendon Press, Oxford (1987); Singh, D.P., *Breeding for*

- 63 -

Resistance to Diseases and Insect Pests, Springer-Verlag, NY (1986); and Wricke and Weber, Quantitative Genetics and Selection Plant Breeding, Walter de Gruyter and Co., Berlin (1986).

The genetic properties engineered into the transgenic seeds and plants described above are passed on by sexual reproduction or vegetative growth and can thus be maintained and propagated in progeny plants. Generally said maintenance and propagation make use of known agricultural methods developed to fit specific purposes such as tilling, sowing or harvesting. Specialized processes such as hydroponics or greenhouse technologies can also be applied. As the growing crop is vulnerable to attack and damages caused by insects or infections as well as to competition by weed plants, measures are undertaken to control weeds, plant diseases, insects, nematodes, and other adverse conditions to improve yield. These include mechanical measures such a tillage of the soil or removal of weeds and infected plants, as well as the application of agrochemicals such as herbicides, fungicides, gametocides, nematicides, growth regulants, ripening agents and insecticides.

Use of the advantageous genetic properties of the transgenic plants and seeds according to the invention can further be made in plant breeding, which aims at the development of plants with improved properties such as tolerance of pests, herbicides, or stress, improved nutritional value, increased yield, or improved structure causing less loss from lodging or shattering. The various breeding steps are characterized by well-defined human intervention such as selecting the lines to be crossed, directing pollination of the parental lines, or selecting appropriate progeny plants. Depending on the desired properties, different breeding measures are taken. The relevant techniques are well known in the art and include but are not limited to hybridization, inbreeding, backcross breeding, multiline breeding, variety blend, interspecific hybridization, aneuploid techniques, etc. Hybridization techniques also include the sterilization of plants to yield male or female sterile plants by mechanical, chemical, or biochemical means. Cross pollination of a male sterile plant with pollen of a different line assures that the genome of the male sterile but female fertile plant will uniformly obtain properties of both parental lines. Thus, the transgenic seeds and plants according to the invention can be used for the breeding of improved plant lines, that for example, increase the effectiveness of conventional methods such as herbicide or pestidice treatment or allow one to dispense with said methods due to their modified genetic properties. Alternatively new crops with improved stress tolerance can be obtained, which, due to their optimized genetic "equipment", yield harvested product of - 64 -

better quality than products that were not able to tolerate comparable adverse developmental conditions.

Example 27: Seed Production

In seed production, germination quality and uniformity of seeds are essential product characteristics, whereas germination quality and uniformity of seeds harvested and sold by the farmer is not important. As it is difficult to keep a crop free from other crop and weed seeds, to control seedborne diseases, and to produce seed with good germination, fairly extensive and well-defined seed production practices have been developed by seed producers, who are experienced in the art of growing, conditioning and marketing of pure seed. Thus, it is common practice for the farmer to buy certified seed meeting specific quality standards instead of using seed harvested from his own crop. Propagation material to be used as seeds is customarily treated with a protectant coating comprising herbicides, insecticides, fungicides, bactericides, nematicides, molluscicides, or mixtures thereof. Customarily used protectant coatings comprise compounds such as captan, carboxin, thiram (TMTD*), methalaxyl (Apron*), and pirimiphos-methyl (Actellic*). If desired, these compounds are formulated together with further carriers, surfactants or applicationpromoting adjuvants customarily employed in the art of formulation to provide protection against damage caused by bacterial, fungal or animal pests. The protectant coatings may be applied by impregnating propagation material with a liquid formulation or by coating with a combined wet or dry formulation. Other methods of application are also possible such as treatment directed at the buds or the fruit.

It is a further aspect of the present invention to provide new agricultural methods, such as the methods examplified above, which are characterized by the use of transgenic plants, transgenic plant material, or transgenic seed according to the present invention.

The seeds may be provided in a bag, container or vessel comprised of a suitable packaging material, the bag or container capable of being closed to contain seeds. The bag, container or vessel may be designed for either short term or long term storage, or both, of the seed. Examples of a suitable packaging material include paper, such as kraft paper, rigid or pliable plastic or other polymeric material, glass or metal. Desirably the bag, container, or vessel is comprised of a plurality of layers of packaging materials, of the same or differing type. In one embodiment the bag, container or vessel is provided so as to

exclude or limit water and moisture from contacting the seed. In one example, the bag, container or vessel is sealed, for example heat sealed, to prevent water or moisture from entering. In another embodiment water absorbent materials are placed between or adjacent to packaging material layers. In yet another embodiment the bag, container or vessel, or packaging material of which it is comprised is treated to limit, suppress or prevent disease, contamination or other adverse affects of storage or transport of the seed. An example of such treatment is sterilization, for example by chemical means or by exposure to radiation. Comprised by the present invention is a commercial bag comprising seed of a transgenic plant comprising a gene of the present invention that is expressed in said transformed plant at higher levels than in a wild type plant, together with a suitable carrier, together with label instructions for the use thereof for conferring broad spectrum disease resistance to plants.

BUDAPLET TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT OF MICROORGANISMS FOR THE PURPOSE OF PATENT PROCEDURES

INTERNATIONAL FORM

TO

VIABILITY STATEMENT

Novartie AG Novartie Corporation 3054 Cornwallis Rd. Research Triangle Park, NC 27709

issued pursuant to Rule 10.2 by the INTERNATIONAL DEPOSITARY AUTHORITY identified at the bottom of this page

NAME AND ADDRESS OF THE PARTY TO WHOM THE VIRBILITY STATEMENT IS ISSUED

I. DEPOSITOR	II. IDENTIFICATION OF THE MICROORGANISM		
Name: Novartis AG Novartis Corporation Address: 3054 Cornwallis Rd. Research Triangle Park, NC 27709	Depositor's taxonomic designation and accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY: Escherichia coli NRRL B-30077 Date of: October 28, 1998 X 1 Original Deposit 1 Repropagation of Original Deposit		
III. (a) VIABILITY STATEMENT			
Deposit was found: X Viable Nonviable on October 31, 1998 (Date) International Depositary Authority's preparation was found viable on December 8, 1998 (Date)			
III. (b) DEPOSITOR'S EQUIVALENCY DECL	ARATION		
Depositor determined the International Depositary Authority's preparation was			
Signature of Depositor Not equivalent to deposit on 1-6-99 (Date)			
IV. CONDITIONS UNDER WHICH THE VIABILITY TEST WAS PERFORMED (Depositors/Depositary)			
The dried culture was put into 2 mls LBampusymin and grown at 37°C overnight with shaking. Some of the liquid culture was streaked to an LBampusylliate + grown at 37°C overnight.			
V. INTERNATIONAL DEPOSITARY AUTHORIT	Y		
Name: Agricultural Research Culture Collection (NRRL) International Depositary Authorit	1.2. Las		
Address: 1815 N. University Street Peoris, Illinois 61604 U.S.A.	/2-3-1Y		

Indicate the date of the original deposit or when a new deposit has been made.

Mark with a cross the applicable box.

In the cases referred to in Rule 10.2(a)(ii) and (iii), refer to the most recent viability test.

Fill in if the information has been requested.

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT OF MICROORGANISMS FOR THE PURPOSE OF PATENT PROCEDURES

INTERNATIONAL FORM

TO
Novartis AG
Novartis Corporation
3054 Cornwallis Rd.
Research Triangle Park,
NC 27709

RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT issued pursuant to Rule 7.1 by the INTERNATIONAL DEPOSITARY AUTHORITY identified at the bottom of this page

NAME AND ADDRESS OF DEPOSITOR

I. IDENTIFICATION OF THE MICROGRANISM				
Identification reference given by the DEPOSITOR:	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY:			
Escherichia coli PNOV2400	NRRL 8-30077			
II. SCIENTIFIC DESCRIPTION AND/OR PROPOS	ED TAXONOMIC DESIGNATION			
The microorganism identified under I. above	e was accompanied by:			
a scientific description				
x a proposed taxonomic designation				
(Mark with a cross where soplicable)				
III. RECEIPT AND ACCEPTANCE				
This International Depositary Authority accepts the microorganism identified under I. above, which was received by it on October 28, 1998(date of the original deposit)				
IV. RECEIPT OF REQUEST FOR CONVERSION				
The microorganism identified under I. above was received by this International Depositary Authority on (date of the original deposit) and a request to convert the original deposit to a deposit under the Eudapest Treaty was received by it on (date of receipt of request for conversion).				
V. INTERNATIONAL DEPOSITARY AUTHORITY				
Name: Agricultural Research Culture Collection (NRRL) International Depositary Authority Address: 1815 N. University Street	Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorized official(s):			
Peoria, Illinois 61604 U.S.A.	Date: /2 3-1/			

Where Rule 6.4(d) applies, such date is the date on which the status f international depositary authority was acquired.

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT OF MICROGRAMISMS FOR THE PURPOSE OF PATENT PROCEDURES

INTERNATIONAL PORK

TO
Novertis AG
Novertis Corporation
3054 Cornwallis Rd.
Research Triangle Park,
NC 27709

RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT issued pursuant to Rule 7.1 by the INTERNATIONAL DEPOSITARY AUTHORITY identified at the bottom of this page

NAME AND ADDRESS OF DEPOSITOR

<u> </u>			
I. IDENTIFICATION OF THE MICROORGANISM			
Identification reference given by the DEPOSITOR:	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY:		
Escherichia coli pNOV1001	MRRL 8-30078		
II. SCIENTIFIC DESCRIPTION AND/OR PROPOS	SED TAXONOMIC DESIGNATION		
The microorganism identified under I. about	re was accompanied by:		
a scientific description			
x a proposed taxonomic designation			
(Mark with a cross where applicable)			
III. RECEIPT AND ACCEPTANCE			
This International Depositary Authority accepts the microorganism identified under I. above, which was received by it on October 28, 1998(date of the original deposit)			
IV. RECEIPT OF REQUEST FOR CONVERSION			
The microorganism identified under I. above was received by this International Depositary Authority on (date of the original deposit) and a request to convert the original deposit to a deposit under the Budupest Treaty was received by it on (date of receipt of request for conversion).			
V. INTERNATIONAL DEPOSITARY AUTHORITY			
Name: Agricultural Research Culture Collection (NRRL) International Depositary Authority	Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorized official(s):		
Address: 1815 N. University Street	Date: 12 -21		

Where Rule 6.4(d) applies, such date is the date on which the status of international depositary authority was acquired.

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION P THE DEPOSIT F MICROGREAMISMS FOR THE PURPOSE OF PATENT PROCEDURES

INTERNATIONAL FORM

VIABILITY STATEMENT

Novartis AG Novartis Corporation 3054 Cornwellis Rd. Research Triangle Park, NC 27709

issued pursuant to Rule 10.2 by the INTERNATIONAL DEPOSITARY AUTHORITY identified at the bottom of this page

NAME AND ADDRESS OF THE PARTY TO WHOM TARTITY STATEMENT IS ISSUED

THE VINBILLY BINTERENT TO ISSUED			
I. DEPOSITOR	II. IDENTIFICATION OF THE MICROORGANISM		
Name: Novartis AG Novartis Corporation Address: 3054 Cornwallis Rd. Research Triangle Park, NC 27709	Depositor's taxonomic designation and accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY: **Escherichia coli NRRL B-30078** Date of: October 28, 1998** **Y Original Deposit**		
III. (a) VIRBILITY STATEMENT			
Deposit was found: K Viable Nonviable on October 31, 1998 (Date) International Depositary Authority's preparation was found viable on December 8, 1998(Date)			
III. (b) DEPOSITOR'S EQUIVALENCY DECLARATION			
Depositor determined the International Depositary Authority's preparation was I Equivalent 1 Not equivalent to deposit on 1-6-99 (Date) Signature of Depositor Nope Hant			
IV. CONDITIONS UNDER WHICH THE VIABILITY TEST WAS PERFORMED (Depositors/Depositary)			
The dried culture was put into 2 mls LBamp(100mg/ml) and grown at 37°C overnight with shaking. Some of the liquid culture was streaked to an LBamp plate and grown at 37°C overnight.			
V. INTERNATIONAL DEPOSITARY AUTHORITY			
Name: Agricultural Research Culture Collection (NRRL) International Depositary Authority Address: 1815 N. University Street Peoris. Illinois 61604 U-S.A.	Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorized official(s):		

Indicate the date of the original deposit or when a new deposit has been made.

* Mark with a cross the applicable box.

* In the cases referred to in Rule 10.2(a)(ii) and (iii), refer to the most recent visbility test.

* Fill in if the information has been requested.

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT OF MICROORGANISMS FOR THE PURPOSE OF PATENT PROCEDURES

INTERNATIONAL FORM

TO

Novartis Corp. c/o Novartis AG P. O. Box 12257

Research Triangle Park, NC 27709

RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT issued pursuant to Rule 7.1 by the INTERNATIONAL DEPOSITARY AUTHORITY identified at the bottom of this page

NAME AND ADDRESS OF DEPOSITOR

I. IDENTIFICATION OF THE MICROORGANISM			
Identification reference given by the DEPOSITOR:	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY:		
Bacteria sp. pCIB 9359-7	NRRL B-21835		
II. SCIENTIFIC DESCRIPTION AND/OR PROPOS	ED TAXONOMIC DESIGNATION		
The microorganism identified under I. above	e was accompanied by:		
a scientific description	•		
a proposed taxonomic designation			
(Mark with a cross where applicable)			
III. RECEIFT AND ACCEPTANCE			
This International Depositary Authority accepts the microorganism identified under I. above, which was received by it on September 17, 1997 (date of the original deposit)'			
IV. RECEIPT OF REQUEST FOR CONVERSION			
The microorganism identified under I. above was received by this International Depositary Authority on (date of the original deposit) and a request to convert the original deposit to a deposit under the Budapest Treaty was received by it on (date of receipt of request for conversion).			
V. INTERNATIONAL DEPOSITARY AUTHORITY			
Name: Agricultural Research Culture Collection (NRRL) International Depositary Authority	Signature (s) of person (s) having the power to represent the International Depositary Authority or of authorized official (s):		
Address: 1815 N. University Street Peoria, Illinois 61604 U.S.A.	Date: 11-13-97		

^{&#}x27;Where Rule 6.4(d) applies, such date is the date on which the status of international depositary authority was acquired.

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT OP MICROORGANISMS FOR THE PURPOSE OF PATENT PROCEDURES

INTERNATIONAL FORM

TO

VIABILITY STATEMENT

Novartis Corp. c/o Novartis AG P. O. Box 12257 Research Triangle Park, NC 27709

issued pursuant to Rule 10.2 by the INTERNATIONAL DEPOSITARY AUTHORITY identified at the bottom of this page

NAME AND ADDRESS OF THE PARTY TO WHOM THE VIABILITY STATEMENT IS ISSUED

	II. IDENTIFICATION OF THE MICROORGANISM
I. DEPOSITOR Name: Novartis Corp c/o Novartis AG Address: P. O. Box 12257 Research Triangle Park, NC 27709	Depositor's taximomic designation and accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY: Bacteria sp. NRRL B-21835 Date of:September 17, 1997 2 Original Deposit 2 New Deposit: 2 Repropagation of Original Deposit
III. (a) VIABILITY STATEMENT	
Deposit was found: Viable Nonviab International Depositary Authority's prepa	
III. (b) DEPOSITOR'S EQUIVALENCY DECLARA	ATION
Depositor determined the International Depositor	positary Authority's preparation was
IV. CONDITIONS UNDER WHICH THE VIABILITY	TEST WAS PERFORMED (Depositors/Depositary)'
V. INTERNATIONAL DEPOSITARY AUTHORITY	
Name: Agricultural Research Culture Collection (NRRL) International Depositary Authority Address: 1815 N. University Street	Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorized official(s):
Peoria, Illinois 61604 U.S.A.	Pace: //- 17 . 77

Indicate the date of the original deposit or when a new deposit has been made.

^{*} Mark with a cross the applicable box.

In the cases referred to in Rule 10.2(a)((1) and ((()), refer to the most secure viability test.

Fill in if the information has been requested.

What is claimed is:

- 1. An isolated nucleic acid molecule comprising:
 - (a) a nucleotide sequence substantially similar to a nucleotide sequence selected from the group consisting of: nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, nucleotides 4515-9269 of SEQ ID NO:1, nucleotides 15,171-18,035 of SEQ ID NO:11, and nucleotides 31,393-35,838 of SEQ ID NO:11;
 - (b) a nucleotide sequence comprising nucleotides 23,768-31,336 of SEQ ID NO:11; or
- (c) a nucleotide sequence isocoding with the nucleotide sequence of (a) or (b); wherein expression of said nucleic acid molecule results in at least one toxin that is active against insects.
- 2. An isolated nucleic acid molecule comprising a 20 base pair nucleotide portion identical in sequence to a consecutive 20 base pair nucleotide portion of a nucleotide sequence selected from the group consisting of: nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, nucleotides 4515-9269 of SEQ ID NO:1, nucleotides 15,171-18,035 of SEQ ID NO:11, and nucleotides 31,393-35,838 of SEQ ID NO:11, wherein expression of said nucleic acid molecule results in at least one toxin that is active against insects.
- 3. An isolated nucleic acid molecule comprising a nucleotide sequence from *Photorhabdus luminescens* selected from the group consisting of: nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, nucleotides 4515-9269 of SEQ ID NO:1, nucleotides 66-1898 of SEQ ID NO:11, nucleotides 2416-9909 of SEQ ID NO:11, the complement of nucleotides 2817-3395 of SEQ ID NO:11, nucleotides 9966-14,633 of SEQ ID NO:11, nucleotides 14,699-15,007 of SEQ ID NO:11, nucleotides 15,171-18,035 of SEQ ID NO:11, the complement of nucleotides 17,072-17,398 of SEQ ID NO:11, the complement of nucleotides 19,385-20,116 of SEQ ID NO:11, the complement of nucleotides 20,217-20,963 of SEQ ID NO:11,

the complement of nucleotides 22,172-23,086 of SEQ ID NO:11, nucleotides 23,768-31,336 of SEQ ID NO:11, nucleotides 31,393-35,838 of SEQ ID NO:11, the complement of nucleotides 35,383-35,709 of SEQ ID NO:11, the complement of nucleotides 36,032-36,661 of SEQ ID NO:11, and the complement of nucleotides 36,654-37,781 of SEQ ID NO:11.

- 4. An isolated nucleic acid molecule according to claim 1, wherein said nucleotide sequence is substantially similar to nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, or nucleotides 4515-9269 of SEQ ID NO:1.
- 5. An isolated nucleic acid molecule according to claim 1, wherein said nucleotide sequence encodes an amino acid sequence selected from the group consisting of SEQ ID NOs:2-6.
- 6. An isolated nucleic acid molecule according to claim 1, wherein said nucleotide sequence comprises nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides 2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, or nucleotides 4515-9269 of SEQ ID NO:1.
- 7. An isolated nucleic acid molecule according to claim 1, wherein said nucleotide sequence is substantially similar to nucleotides 15,171-18,035 or 31,393-35,838 of SEQ ID NO:11.
- 8. An isolated nucleic acid molecule according to claim 1, wherein said nucleotide sequence encodes the amino acid sequence set forth in SEQ ID NOs:12-14.
- 9. An isolated nucleic acid molecule according to claim 1, wherein said nucleotide sequence comprises nucleotides 15,171-18,035; 23,768-31,336; or 31,393-35,838 of SEQ ID NO:11.
- 10. An isolated nucleic acid molecule according to claim 2, comprising a 20 base pair nucleotide portion identical in sequence to a consecutive 20 base pair nucleotide portion of nucleotides 412-1665 of SEQ ID NO:1, nucleotides 1686-2447 of SEQ ID NO:1, nucleotides

2758-3318 of SEQ ID NO:1, nucleotides 3342-4118 of SEQ ID NO:1, or nucleotides 4515-9269 of SEQ ID NO:1.

- 11. An isolated nucleic acid molecule according to claim 2, comprising a 20 base pair nucleotide portion identical in sequence to a consecutive 20 base pair nucleotide portion of nucleotides 15,171-18,035 or 31,393-35,838 of SEQ ID NO:11.
- 12. A chimeric gene comprising a heterologous promoter sequence operatively linked to the nucleic acid molecule of claim 1 or claim 2.
- 13. A recombinant vector comprising the chimeric gene of claim 12.
- 14. A host cell comprising the chimeric gene of claim 12.
- 15. A host cell according to claim 14, which is a bacterial cell.
- 16. A-host cell according to claim 14, which is a yeast cell.
- 17. A host cell according to claim 14, which is a plant cell.
- 18. A plant comprising the plant cell of claim 17.
- 19. A plant according to claim 18, which is maize.
- 20. A toxin produced by the expression of a DNA molecule according to claim 1 or claim 2.
- 21. A toxin according to claim 20, wherein said toxin has activity against Lepidopteran insects.
- 22. A toxin according to claim 21, wherein said toxin has activity against *Plutella xylostella* (Diamondback Moth), *Trichoplusia ni* (Cabbage Looper), *Ostrinia nubilalis* (European Corn Borer), *Heliothis virescens* (Tobacco Budworm), *Helicoverpa zea* (Corn Earworm), *Spodoptera exigua* (Beet Armyworm), and *Spodoptera frugiperda* (Fall Armyworm).

- 23. A toxin according to claim 20, wherein said toxin has activity against Lepidopteran and Coleopteran insects.
- 24. A toxin according to claim 23, wherein said toxin has insecticidal activity against Plutella xylostella (Diamondback Moth), Ostrinia nubilalis (European Corn Borer), and Manduca sexta (Tobacco Hornworm), Diabrotica virgifera virgifera (Western Corn Rootworm), Diabrotica undecimpunctata howardi (Southern Corn Rootworm), and Leptinotarsa decimlineata (Colorado Potato Beetle).
- 25. A toxin according to claim 20, wherein said toxin comprises an amino acid sequence selected from the group consisting of: SEQ ID NOs:2-6.
- 26. A toxin according to claim 20, wherein said toxin comprises an amino acid sequence selected from the group consisting of: SEQ ID NOs:12-14.
- 27. A composition comprising an insecticidally effective amount of a toxin according to claim 20.
- 28. A method of producing a toxin that is active against insects, comprising:
 - (a) obtaining the host cell of claim 14; and
 - (b) expressing the nucleic acid molecule in said cell, which results in at least one toxin that is active against insects.
- 29. A method of producing an insect-resistant plant, comprising introducing a nucleic acid molecule according to claim 1 into said plant, wherein said nucleic acid molecule is expressible in said plant in an effective amount to control insects.
- 30. A method of controlling insects comprising delivering to the insects an effective amount of a toxin according to claim 44.
- 31. The method of claim 29 or claim 30, wherein the insects are Lepidopteran insects.

- 32. The method of claim 31, wherein the insects are selected from the group consisting of: Plutella xylostella (Diamondback Moth), Trichoplusia ni (Cabbage Looper), Ostrinia nubilalis (European Corn Borer), Heliothis virescens (Tobacco Budworm), Helicoverpa zea (Corn Earworm), Spodoptera exigua (Beet Armyworm), and Spodoptera frugiperda (Fall Armyworm).
- 33. The method of claim 29 or claim 30, wherein the insects are Lepidopteran and Coleopteran insects.
- 34. The method of claim 33, wherein the insects are selected from the group consisting of: *Plutella xylostella* (Diamondback Moth), *Ostrinia nubilalis* (European Corn Borer), and *Manduca sexta* (Tobacco Hornworm), *Diabrotica virgifera virgifera* (Western Corn Rootworm), *Diabrotica undecimpunctata howardi* (Southern Corn Rootworm), and *Leptinotarsa decimlineata* (Colorado Potato Beetle).
- 35. The method of claim 30, wherein the toxin is delivered to the insects orally.
- 36. A method for mutagenizing a nucleic acid molecule according to claim 1, wherein the nucleic acid molecule has been cleaved into population of double-stranded random fragments of a desired size, comprising:
 - (a) adding to the population of double-stranded random fragments one or more single- or double-stranded oligonucleotides, wherein said oligonucleotides each comprise an area of identity and an area of heterology to a doublestranded template polynucleotide;
 - (b) denaturing the resultant mixture of double-stranded random fragments and oligonucleotides into single-stranded fragments;
 - (c) incubating the resultant population of single-stranded fragments with a polymerase under conditions which result in the annealing of said single-stranded fragments at said areas of identity to form pairs of annealed fragments, said areas of identity being sufficient for one member of a pair to prime replication of the other, thereby forming a mutagenized double-stranded polynucleotide; and

(d) repeating the second and third steps for at least two further cycles, wherein the resultant mixture in the second step of a further cycle includes the mutagenized double-stranded polynucleotide from the third step of the previous cycle, and wherein the further cycle forms a further mutagenized double-stranded polynucleotide.

```
SEQUENCE LISTING
<110> Novartis AG
<120> Novel Toxins And Uses Thereof
<130> PI/5-30421/A/CGC 1963
<140>
<141>
<160> 22
<170> PatentIn Ver. 2.0
<210> 1
<211> 9717
<212> DNA
<213> Photorhabdus luminescens
<220>
<221> CDS
<222> (412)..(1665)
<223> orf1 ~46.4 kDa
<220>
<221> CDS
<222> (1686)..(2447)
<223> orf2 ~28.1kDa
<220>
<221> CDS
<222> (2758)..(3318)
<223> orf3 ~20.7 kDa
<220>
<221> CDS
<222> (3342)..(4118)
<223> orf4 ~28.7 kDa
<220>
<221> CDS
<222> (4515)..(9269)
<223> orf5 ~176 kDa
<400> 1
gaattcatat gctatgaaat aaacagttgg cgcaataatt aaagctatta tttttatttt 60
gtttttatac aatgatatgc tttattaaac agaataatga gttaatgata aataaatcct 120
cgggatttat catgatatta tggccgaatg tgatgtgaac aattatttta taattagatt 180
aataatataa tggtattaaa ataacaatat atttattcat gggtatttat catcggtttt 240
attacatggg gaataatcta taaattagtt ttacataatt cacaaatagc gattccatta 300
 accaggaata ttaaaaatac ttatttatga ttatggtgat atatcttcat tagcctactt 360
 ttataactag aaaaattgac attttcaatc catgtataaa tggtaaccaa t atg cag 417
                                                          Met Gln
 aga get caa ega gtt gtt att aca ggt atg ggt gee gta aca eeg att
 Arg Ala Gln Arg Val Val Ile Thr Gly Met Gly Ala Val Thr Pro Ile
```

ggt Gly	gaa Glu 20	gat Asp	gtt Val	gaa Glu	tca Ser	Cys 25	tgg Trp	Gln	agt Ser	Ile	Ile 30	Glu	aaa Lys	caa Gln	His	212
cga Arg 35	ttt Phe	cac His	aga Arg	att Ile	gaa Glu 40	ttt Phe	cct Pro	gac Asp	tca Ser	ttc Phe 45	att Ile	aat Asn	tcg Ser	cgt Arg	ttc Phe 50	561
ttt Phe	tct Ser	ttc Phe	ctt Leu	gca Ala 55	cca Pro	aac Asn	cca Pro	tcc Ser	cgc Arg 60	tat Tyr	cag Gln	tta Leu	tta Leu	cca Pro 65	aaa Lys	609
aag Lys	ttg Leu	act Thr	cat His 70	aca Thr	ctt Leu	tct Ser	gac Asp	tgc Cys 75	gga Gly	aaa Lys	gca Ala	gca Ala	ttg Leu 80	aag Lys	gcg Ala	657
act Thr	tat Tyr	caa Gln 85	gct Ala	ttt Phe	acc Thr	caa Gln	gca Ala 90	ttc Phe	ggc Gly	gtg Val	aat Asn	ata Ile 95	tca Ser	cct Pro	gtt Val	705
gaa Glu	tat Tyr 100	tac Tyr	gat Asp	aaa Lys	tac Tyr	gaa Glu 105	tgt Cys	ggc Gly	gta Val	att Ile	ctt Leu 110	ggc Gly	agt Ser	ggt Gly	tgg Trp	753
gga Gly 115	gct Ala	att Ile	gat Asp	aat Asn	gcc Ala 120	gga Gly	gat Asp	cat His	gct Ala	tgc Cys 125	caa Gln	tat Tyr	aag Lys	caa Gln	gca Ala 130	801
aaa Lys	tta Leu	gct Ala	cat His	cct Pro 135	atg Met	agt Ser	aat Asn	ctt Leu	att Ile 140	acc Thr	atg Met	cca Pro	agc Ser	tcc Ser 145	atg Met	849
				Ser				gga Gly 155								897
			Ala					Thr					Asp		ttt Phe	945
		Ile										Ala			gct Ala	993
gaa Glu 195	Ser	cto Leu	acg Thr	cgg Arg	gaa Glu 200	Cys	aat Asr	att Ile	tgg Trp	agt Ser 205	Ile	gat Asp	gta Val	ctg Leu	aat Asn 210	1041
					Gln					Leu					ttt Phe	1089
				ser Ser					Ala					Va]	gtt Val	1137
			ı Asr					ı Ile					Thr		tta Leu	1185
gcg	gaa Glu 260	ı Ile	aaa Lys	a ggt s Gly	tac Tyn	gcc Ala 265	Gl	a tat n Tyn	tca Ser	a gat Asp	gco Ala 270	a Val	aat Asr	tta Leu	a acc ı Thr	1233

egg Arg 275	Pro	aca Thr	gaa Glu	gat Asp	att Ile 280	gaa Glu	cct Pro	aaa Lys	ata Ile	tta Leu 285	gcg Ala	ata Ile	act Thr	aaa Lys	gcc Ala 290	1281
att Ile	gag Glu	cag Gln	gca Ala	cag Gln 295	att Ile	tcg Ser	ccg Pro	aaa Lys	gat Asp 300	att Ile	gac Asp	tac Tyr	att Ile	aat Asn 305	act	1329
cat His	ggt Gly	act Thr	tct Ser 310	aca Thr	ccg Pro	tta Leu	aat Asn	gat Asp 315	ctt Leu	tat Tyr	gaa Glu	act Thr	cag Gln 320	gca Ala	att Ile	1377
aaa Lys	gca Ala	gca Ala 325	ctg Leu	ggc Gly	caa Gln	tat Tyr	gct Ala 330	tat Tyr	cag Gln	gta Val	cct Pro	ata Ile 335	tca Ser	agc Ser	aca Thr	1425
aaa Lys	tct Ser 340	tat Tyr	acc Thr	ggc Gly	cac His	ctt Leu 345	att Ile	gct Ala	gcc Ala	gcc Ala	ggt Gly 350	agt Ser	ttt Phe	gaa Glu	acg Thr	1473
att Ile 355	Val	tgt Cys	gtg Val	aaa Lys	gca Ala 360	tta Leu	gct Ala	gaa Glu	aat Asn	tgc Cys 365	ttg Leu	cca Pro	gca Ala	aca Thr	ttg Leu 370	1521
aat Asn	tta Leu	cac His	cgg Arg	gcc Ala 375	gat Asp	cca Pro	gat. Asp	tgc Cys	gat Asp 380	ctc Leu	aat Asn	tat Tyr	ttg Leu	cct Pro 385	aat Asn	1569
caa Gln	cat His	tgc Cys	tac Tyr 390	acc Thr	gct Ala	caa Gln	cca Pro	gag Glu 395	gtg Val	aca Thr	ctc Leu	aat Asn	att Ile 400	agc Ser	gca Ala	1617
ggt Gly	ttc Phe	ggc Gly 405	Gly ggg	cat His	aac Asn	gct Ala	gcg Ala 410	ttg Leu	gtt Val	atc Ile	gct Ala	aag Lys 415	gta Val	agg Arg	taa	1665
GIY	Pne	G1y 405	GIY	His	Asn ca at	Ala Eg ga	Ala 410 aa ga lu As	Leu at at	Val :t ga	Ile a ca	Ala at to	Lys 415 g to p Se	Val eg aa	Arg	taa c tct ne Ser	
ggg ctg	Phe atato gat	Gly 405 gtt g	Gly gattt aac	His	Asn ca at Me	Ala ig ga et Gl 42 cat	Ala 410 aa ga lu As 20 tat	Leu at at ap Il	Val t ga le Gl	Ile a ca u Hi aaa	Ala at to is Tr 42 age	Lys 415 gg to op Se 25	Val eg aæ er As	Arg at tt sn Ph	cc tct ne Ser	
ctg ctg ggg Gly 430	Phe atato gat Asp ata	Gly 405 gtt c ttt Phe cag	gatti aac Asn caa	His ttg:	Asn Ca at Me atc Ile 435	Ala tg ga et Gl 42 cat His	Ala 410 aa ga lu As 20 tat Tyr	Leu at at sp II tcg Ser	Val t ga le Gl gcg Ala atg	Ile aa ca u Hi aaa Lys 440	Ala at to is Th 42 agc Ser	Lys 415 gg to to Se 25 gag Glu	Val eg aæ er As tct Ser	Arg at tt en Pl ttg Leu	cgc Arg 445	1718
ctg. ggg Gly 430 aat Asn	gat Asp ata Ile	gtt g ttt Phe cag Gln	gatti aac Asn caa Gln	ccc Pro	Asn Ca at Me atc Ile 435 Ccg Pro	Ala cg ga et Gl 42 cat His gtg Val	Ala 410 aa ga lu As 20 tat Tyr cag Gln	Leu at at sp II tcg Ser gga Gly	Val t ga te Gl gcg Ala atg Met 455	Ile aa ca u Hi aaa Lys 440 ttg Leu	Ala at to is Tr 42 agc ser agt Ser	Lys 415 gg to p Se 25 gag Glu ttg Leu	Val eg aæ er As tet Ser etc Leu	Arg at tt sn Pl ttg Leu tat Tyr 460	cgc tct cgc Arg 445 gta Val	1718 1766
ggg Gly 430 aat Asn cgg Arg	gat Asp ata Ile caa Gln att	ttt Phe cag Gln cag Gln gat	aac Asn caa Gln ttt Phe 465 gcc	ccc Pro cac His 450	Asn ca at Me atc Ile 435 ccg Pro caa Gln	Ala cg ga et Gi 42 cat His gtg Val tta Leu cgc	Ala 410 410 tat Tyr cag Gln act Thr	Leu tcg Ser gga Gly tcc Ser 470	Val tt ga le Gl gcg Ala atg Met 455 gct Ala	Ile Ile aaa ca u H aaaa Lys 440 ttg Leu ttt	Ala at to is Th 42 agc Ser agt Ser aca Thr	Lys 415 gg to p Se 25 gag Glu ttg Leu acg Thr	Val eg aæ er As tet Ser etc Leu gga Gly 475	Arg at the trian Pl ttg Leu tat Tyr 460 ata Ile	cgc Arg 445 gta Val ttg Leu	1718 1766 1814
ggg Gly 430 aat Asn cgg Arg	gat Asp ata Ile caa Gln att Ile	ttt Phe cag Gln cag Gln gat Asp 480 agg	aac Asn caa Gln ttt Phe 465 gcc Ala att	ccc Pro cac His 450 tct Ser	Asn ca at Me atc Ile 435 ccg Pro caa Gln ttc Phe act	Ala cg ga et Gl 42 cat His gtg Val tta Leu cgc Arg	Ala 410 Ala	Leu tcg Ser gga Gly tcc Ser 470 tat Tyr	Val tt ga le Gl gcg Ala atg Met 455 gct Ala gtt Val	Ile Ile Ile Ile Ile Ile Ile Itt Ile	Ala at to is Tr 42 agc Ser agt Ser aca Thr acc Thr	Lys 415 gg to p Se 25 gag gag Glu ttg Leu acg Thr gca Ala 490	Val cg aa tct Ser ctc Leu gga Gly 475 tta Leu	Arg at the sn Ph ttg Leu tat Tyr 460 ata Ile ccc Pro	cgc Arg 445 gta Val ttg Leu cat	1718 1766 1814 1862
ggg Gly 430 aat Asn cgg Arg aac Asn	gat Asp ata Ile caa Gln att Ile ctg Leu 495 aaa	Gly 405 ttt Phe cag Gln cag Gln gat Asp 480 agg Arg	aac Asn caa Gln ttt Phe 465 gcc Ala att Ile	ccc Pro cac His 450 tct Ser tct Ser	Asn Ca at Me atc Ile 435 ccg Pro caa Gln ttc Phe act act ttg	Ala cg ga et Gl 42 cat His gtg Val tta Leu cgc Arg aaa Lys 500 ttc	Ala 410 Ala 420 tat Tyr cag Gln act Thr cag Sln 485 aac Asn	Leu tcg Ser gga Gly tcc Ser 470 tat Tyr aaaa Lys	Val t gate Gl gcg Ala atg Met 455 gct Ala gtt Val acg Thr	Ile Ile Ile Ile Ile Ile Ilys Ile Ile Itt Ile	Ala At too	Lys 415 gg tc pp Se 25 gag Glu ttg Leu acg Thr gca Ala 490 tta Leu	Val cg aa cg aa tct Ser ctc Leu gga Gly 475 tta Leu gaa Glu aat	Arg Arg At the series of the	cgc tct ne Ser cgc Arg 445 gta Val ttg Leu cat His	1718 1766 1814 1862 1910

Glu	Ser	Ile	Glu	Asp 530	Trp	Ile	Val	Gln	Asp 535	Asn	Cys	Gln	Lys	Leu 540	Thr	
ata Ile	aca Thr	Gly ggg	gag Glu 545	gaa Glu	gtt Val	tgt Cys	gaa Glu	aag Lys 550	tat Tyr	gct Ala	gtc Val	ttt Phe	aga Arg 555	tac Tyr	tat Tyr	2102
ttc Phe	cca Pro	agt Ser 560	gtc Val	act Thr	tct Ser	att Ile	gga Gly 565	tgg Trp	ttc Phe	ctg Leu	gat Asp	gcg Ala 570	ctt Leu	gct Ala	ttt Phe	2150
cat His	ctt Leu 575	att Ile	att Ile	aat Asn	tcg Ser	aca Thr 580	gga Gly	ttt Phe	ctt Leu	aat Asn	ttt Phe 585	gag Glu	cac His	tac Tyr	cat His	2198
ttt Phe 590	aac Asn	caa Gln	tta Leu	cag Gln	gat Asp 595	tat Tyr	ctg Leu	agt Ser	caa Gln	tct Ser 600	ttt Phe	act Thr	ttg Leu	cat His	act Thr 605	2246
Gly ggg	caa Gln	gcg Ala	att Ile	aaa Lys 610	atc Ile	agg Arg	aag Lys	gag Glu	att Ile 615	gtt Val	aat Asn	agt Ser	aca Thr	gta Val 620	tta Leu	2294
tta Leu	tct Ser	tca Ser	ccg Pro 625	gat Asp	atc Ile	tgt Cys	gtt Val	gaa Glu 630	tta Leu	aat Asn	cct Pro	cct Pro	tta Leu 635	ttg Leu	att Ile	2342
aag Lys	aat Asn	ggc Gly 640	gat Asp	aaa Lys	gat Asp	tat Tyr	att Ile 645	cgt Arg	att Ile	ttc Phe	tat Tyr	tat Tyr 650	cga Arg	tgt Cys	tta Leu	2390
tat Tyr	gat Asp 655	aaa Lys	aaa Lys	cct Pro	att Ile	ttt Phe 660	gta Val	tca Ser	aag Lys	act Thr	tca Ser 665	att Ile	atc Ile	tct Ser	aag Lys	2438
	aaa Lys	taa	aagg	gaaaq	gcg a	aaat	gccaa	ac ac	caaaq	gtgat	ati	itte	actg			2487
aaat	caaaq	gaa t	tagaa	atati	a at	gato	gaagg	g ata	ataga	aga	tgaa	agaaa	ata a	acaco	cagagt	2547
cct	cttt	.gt. t	tcg	cttga	aa tt	tgat	agto	c ttg	gacta	atgt	ggaa	atco	caa g	gttt	tgtgt	2607
tgga	aagco	gta t	tggta	attgi	cg ct	taaa	agcc	g aad	cttt	ttc	aaat	catt	cct a	attto	caacat	2667
taaa	atgag	gct (cacto	gacta	at tt	caaaa	atcaa	a aat	tgta	atc	tgaa	attt	ta d	cttaa	attatg	2727
ttt	ttca	acc a	attaa	acatı	ta ag	gaggt	tata	a ato Met	gaad Ast	gtt n Val 679	l Lei	a gaz ı Glu	a caa ı Glr	a ggt n Gly	aag Lys 680	2781
gtt Val	gct Ala	gct Ala	tta Leu	tat Tyr 685	tca Ser	gcc Ala	tat Tyr	tcg Ser	gaa Glu 690	aca Thr	gaa Glu	ggt Gly	tct Ser	tcg Ser 695	tgg Trp	2829
gtg Val	gga Gly	aac Asn	ttg Leu 700	tgc Cys	tgt Cys	ttt Phe	tca Ser	agt Ser 705	gat Asp	cgg Arg	gag Glu	cat His	ttg Leu 710	cct Pro	att Ile	2877
atc Ile	gtg Val	aat Asn 715	ggg Gly	cgt Arg	cgt Arg	ttc Phe	ttg Leu 720	att Ile	gaa Glu	ttt Phe	gtt Val	att Ile 725	cca Pro	gat Asp	cat His	2925
tta Leu	ctt Leu	gat Asp	aaa Lys	acg Thr	gtt Val	aaa Lys	ccc Pro	aga Arg	gta Val	ttc Phe	gat Asp	ttg Leu	gat Asp	atc Ile	aat Asn	2973

730	73	5	740)	
aaa caa ttt tta Lys Gln Phe Leu 745					
tta ggt gaa gga Leu Gly Glu Gly					ı Leu
ttc gag tta aat Phe Glu Leu Asn 780					
gct ctt ggt aaa Ala Leu Gly Lys 795	tat gtt gc Tyr Val Al	t att aat a Ile Asn 800	cct tca act Pro Ser Thr	t acg caa tti r Thr Gln Pho 805	atc 3165
ttc ttt gca caa Phe Phe Ala Gln 810		r Ser Glu		t Asn Ala Le	
aca gtt gaa gac Thr Val Glu Asp 825	gaa tta tc Glu Leu Se 830	a aaa cgt r Lys Arg	tat cga gto Tyr Arg Val 835	c aga att at l Arg Ile Ile	cct 3261 Pro 840
gaa ttg caa ggg Glu Leu Gln Gly	ccg tat ta Pro Tyr Ty 845	t ggc ttt r Gly Phe	gaa ctt gat Glu Leu Asp 850	t att ctt tc p Ile Leu Se: 85!	: Ile
aca gct taa tto Thr Ala	acaatat tat	ggagagt gt	Met Glu I	Lys Lys Ile '	Thr Thr
			860	,	365
ttt acc att gag Phe Thr Ile Glu 870	Lys Thr As	t gac aat p Asp Asn 875	ttt tat gct	t aat ggg cg	cat 3410
Phe Thr Ile Glu	Lys Thr As	p Asp Asn 875 t gta ctt	ttt tat gct Phe Tyr Ala aaa caa gaa	t aat ggg cg a Asn Gly Ar 880 a tat agg aa	cat 3410 g His
Phe Thr Ile Glu 870 caa tgt atg gta Gln Cys Met Val	Lys Thr As aaa atc to Lys Ile Se tta gca ct	p Asp Asn 875 t gta ctt r Val Leu 890 t agt gag u Ser Glu	ttt tat got Phe Tyr Ala aaa caa gaa Lys Gln Glu got gaa aaa	t aat ggg cg a Asn Gly Ar 880 a tat agg aa u Tyr Arg As 895 a aga tcg at s Arg Ser Il	cat 3410 g His ggt 3458 n Gly
Phe Thr Ile Glu 870 caa tgt atg gta Gln Cys Met Val 885 gat tgg ata aaa Asp Trp Ile Lys	Lys Thr As aaa atc to Lys Ile Se tta gca ct Leu Ala Le 90 agt gat ag	p Asp Asn 875 t gta ctt r Val Leu 890 t agt gag u Ser Glu c ctc ata	ttt tat gct Phe Tyr Ala aaa caa gaa Lys Gln Glu gct gaa aaa Ala Glu Lys 910 tat gac caa	t aat ggg cg a Asn Gly An 880 a tat agg aa u Tyr Arg As: 895 a aga tcg at s Arg Ser Il	c cat 3410 g His 3458 h Gly 3506 e Gln g cct 3554
Phe Thr Ile Glu 870 caa tgt atg gta Gln Cys Met Val 885 gat tgg ata aaa Asp Trp Ile Lys 900 gtg gcg gca tta Val Ala Ala Leu	Lys Thr As aaa atc to Lys Ile Se tta gca ct Leu Ala Le 90 agt gat ag Ser Asp Se 920 acg aca ga	p Asp Asn 875 t gta ctt r Val Leu 890 t agt gag u Ser Glu 5 c ctc ata r Leu Ile t gca aga	ttt tat gct Phe Tyr Ala aaa caa gaa Lys Gln Glu gct gaa aaa Ala Glu Lys 910 tat gac caa Tyr Asp Glr 925 aat aaa ttt	t aat ggg cg a Asn Gly An 880 a tat agg aa u Tyr Arg As: 895 a aga tcg at s Arg Ser II: 0 a tta aaa at n Leu Lys Me	c cat 3410 g His 3458 h Gly 3506 e Gln 3554 t Pro 930 g tta 3602 y Leu 3410
Phe Thr Ile Glu 870 caa tgt atg gta Gln Cys Met Val 885 gat tgg ata aaa Asp Trp Ile Lys 900 gtg gcg gca tta Val Ala Ala Leu 915 tca ggt tgg aca	Lys Thr As aaa atc to Lys Ile Se tta gca ct Leu Ala Le 90 agt gat ag Ser Asp Se 920 acg aca ga Thr Thr As 935 tat cat gc Tyr His Al	p Asp Asn 875 t gta ctt r Val Leu 890 t agt gag u Ser Glu 5 c ctc ata r Leu Ile t gca aga p Ala Arg	ttt tat gct Phe Tyr Ala aaa caa gaa Lys Gln Glu gct gaa aaa Ala Glu Lys 910 tat gac caa Tyr Asp Glr 925 aat aaa ttt Asn Lys Phe 940 ttt att gac	t aat ggg cg a Asn Gly An 880 a tat agg aa u Tyr Arg Ass 895 a aga tcg at s Arg Ser II. 0 a tta aaa at n Leu Lys Me t gat ctt gg e Asp Leu Gl 94 c qaa cag gt	c cat 3410 g His 3458 h Gly 3506 e Gln 3554 t Pro 930 g tta 3602 y Leu 5 a aca 3650
caa tgt atg gta Gln Cys Met Val 885 gat tgg ata aaa Asp Trp Ile Lys 900 gtg gcg gca tta Val Ala Ala Leu 915 tca ggt tgg aca Ser Gly Trp Thr tta aat ggt gtt Leu Asn Gly Val	Lys Thr As aaa atc to Lys Ile Se tta gca ct Leu Ala Le 90 agt gat ag Ser Asp Se 920 acg aca ga Thr Thr As 935 tat cat gc Tyr His Al	p Asp Asn 875 t gta ctt val Leu 890 t agt gag u Ser Glu 5 c ctc ata Ile t gca aga p Ala Arg t gat gct a Asp Ala 955 c aca aat	ttt tat gct Phe Tyr Ala aaa caa gaa Lys Gln Glu gct gaa aaa Ala Glu Lys 910 tat gac caa Tyr Asp Glr 925 aat aaa ttt Asn Lys Phe 940 ttt att gac Phe Ile Asp	t aat ggg cg a Asn Gly An 880 a tat agg aa u Tyr Arg As: 895 a aga tcg at s Arg Ser Il- 0 a tta aaa at n Leu Lys Me t gat ctt gg e Asp Leu Gl: 94 c gaa cag gt p Glu Gln Va 960 t cag aac ag	c cat 3410 g His 3458 h Gly 3506 e Gln 3554 pro 930 g tta 3602 g tta 3650 l Thr

agc aca gaa tac cta atg gca aaa atg aca ttt gaa gat acg gat ggg Ser Thr Glu Tyr Leu Met Ala Lys Met Thr Phe Glu Asp Thr Asp Gly 995 1000 1005 1010	3794
aaa cgc aca tta aca acg aat atg tca gtt ggt gat gaa gtt ttt gac Lys Arg Thr Leu Thr Thr Asn Met Ser Val Gly Asp Glu Val Phe Asp 1015 1020 1025	3842
agc aag gtt tta tta aaa gcc att gct cct tat gca att aat aca aat Ser Lys Val Leu Leu Lys Ala Ile Ala Pro Tyr Ala Ile Asn Thr Asn 1030 1035 1040	3890
caa ttg cat gaa aac atc aat aca ttg ttt gat aaa aca gaa gag ccg Gln Leu His Glu Asn Ile Asn Thr Leu Phe Asp Lys Thr Glu Glu Pro 1045 1050 1055	3938
aca aaa tcc gat act cat cat caa ata att aat ctt tat cgc tgg aca Thr Lys Ser Asp Thr His His Gln Ile Ile Asn Leu Tyr Arg Trp Thr 1060 1065 1070	3986
ttg cca tat cat ttg agg att ctt gaa ggg aat gac agt act gtt aat Leu Pro Tyr His Leu Arg Ile Leu Glu Gly Asn Asp Ser Thr Val Asn 1075 1080 1085 1090	4034
aga ata tat gtc ctt ggt aaa gag cca tca aat gat aga ttc ctg aca Arg Ile Tyr Val Leu Gly Lys Glu Pro Ser Asn Asp Arg Phe Leu Thr 1095 1100 1105	4082
aga gga agg gta ttt aaa cga gga act cat atg tga atgcacgtga Arg Gly Arg Val Phe Lys Arg Gly Thr His Met 1110 1115	4128
taatgtgagt ggaggatgtg ttatggacta tgcttatacc gtaactattc cggacacgca	4188
gettgetget gaagtgette atgtgacagg gtgttegtgg acgagtggtt attatgatgg	4248
atatcatgat gtcacaatca ttgataacta c ggttgtcag cataaattta gaatttcttc	4308
ggttaatatt ggacgtgcgc taagcatagc gagaataagt tgattttcct tagtaaaaaa	4368
cetttgttta tgctggtaaa cgcatgtgcg tttgccagca attaatatat tecattattg	4428
aaataggaat atagccatat ctgtaattat acataaacga atttttactc gaatataatt	4488
ttaattgatc aaacaggaaa tttaaa atg aaa gct acc gat ata tat tcc aat Met Lys Ala Thr Asp Ile Tyr Ser Asn 1120 1125	4541
gct ttt aat ttc ggt tct tat att aat act ggt gtc gat ccc aga aca Ala Phe Asn Phe Gly Ser Tyr Ile Asn Thr Gly Val Asp Pro Arg Thr 1130 1135 1140	4589
ggt caa tat agt gca aat att aat att atc acg tta aga cct aat aat Gly Gln Tyr Ser Ala Asn Ile Asn Ile Ile Thr Leu Arg Pro Asn Asn 1145 1150 1155	4637
gtg ggt aat tcg gaa caa aca ttg agc cta tca ttc tcg cca tta aca Val Gly Asn Ser Glu Gln Thr Leu Ser Leu Ser Phe Ser Pro Leu Thr 1160 1165 1170 1175	4685
acg tta aac aat ggc ttt ggt att ggc tgg cgc ttt tca tta aca aca Thr Leu Asn Asn Gly Phe Gly Ile Gly Trp Arg Phe Ser Leu Thr Thr 1180 1185 1190	4733

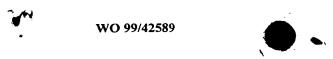
tta gat ata aaa aca ctt aca ttt agc cga gca aat ggg gag caa ttt Leu Asp Ile Lys Thr Leu Thr Phe Ser Arg Ala Asn Gly Glu Gln Phe 1195 1200 1205	4781
aaa tgt aag cca ttg ccg cct aat aat aat gat ctt agt ttt aaa gat Lys Cys Lys Pro Leu Pro Pro Asn Asn Asn Asp Leu Ser Phe Lys Asp 1210 1215 1220	4829
aaa aaa cta aaa gat ttg cgc gta tat aag ctc gat agc aat act ttt Lys Lys Leu Lys Asp Leu Arg Val Tyr Lys Leu Asp Ser Asn Thr Phe 1225 1230 1235	4877
tat gtt tat aac aaa aac ggc att ata gag ata ctt aaa cga att ggg Tyr Val Tyr Asn Lys Asn Gly Ile Ile Glu Ile Leu Lys Arg Ile Gly 1240 1245 1250 1255	4925
tcg agt gat att gca aaa aca gtt gca ctt gaa ttt cct gat ggt gaa Ser Ser Asp Ile Ala Lys Thr Val Ala Leu Glu Phe Pro Asp Gly Glu 1260 1265 1270	4973
gca ttt gat tta att tat aat tca aga ttt gca ttg tcc gaa ata aaa Ala Phe Asp Leu Ile Tyr Asn Ser Arg Phe Ala Leu Ser Glu Ile Lys 1275 1280 1285	5021
tac cgt gtg aca ggt aaa act tat ctt aaa ctc aat tac tct gga aat Tyr Arg Val Thr Gly Lys Thr Tyr Leu Lys Leu Asn Tyr Ser Gly Asn 1290 1295 1300	
aac tgt aca tca gtg gaa tac cct gat gat aat aat att tct gcg aaa Asn Cys Thr Ser Val Glu Tyr Pro Asp Asp Asn Asn Ile Ser Ala Lys 1305 1310 1315	5117
ata gca ttc gat tat cgt aac gat tac ctt att acg gtg act gta cct Ile Ala Phe Asp Tyr Arg Asn Asp Tyr Leu Ile Thr Val Thr Val Pro 1320 1325 1330 1335	•
tac gat gct tct ggt cct att gat tct gcc cga ttt aag atg acc tat Tyr Asp Ala Ser Gly Pro Ile Asp Ser Ala Arg Phe Lys Met Thr Tyr 1340 1345 1350	5213
cag aca tta aaa ggc gta ttt cca gtt atc agc acc ttc cgt aca cca Gln Thr Leu Lys Gly Val Phe Pro Val Ile Ser Thr Phe Arg Thr Pro 1355 1360 1365	
acc ggt tat gtt gag ctg gtg agt tat aaa gag aat ggg cat aaa gtg Thr Gly Tyr Val Glu Leu Val Ser Tyr Lys Glu Asn Gly His Lys Val 1370 1375 1380	
acg gac acg gaa tat att cct tat gcg gct gca ctc act att caa ccc Thr Asp Thr Glu Tyr Ile Pro Tyr Ala Ala Ala Leu Thr Ile Gln Pro 1385 1390 1395	
ggc aat gga caa cct gcg gtc agc aaa tcc tat gaa tat agt tca gta Gly Asn Gly Gln Pro Ala Val Ser Lys Ser Tyr Glu Tyr Ser Ser Val 1400 1405 1410 1419	l
cat aac ttc ttg ggc tat tct tct ggc cgg aca agc ttt gat tcc ag His Asn Phe Leu Gly Tyr Ser Ser Gly Arg Thr Ser Phe Asp Ser Ser 1420 1425 1430	
caa gat aat ttg tat ttg gtc aca ggg aaa tac act tat tca tcc at: Gln Asp Asn Leu Tyr Leu Val Thr Gly Lys Tyr Thr Tyr Ser Ser Il 1435 1440 1445	
gaa cgg gtt tta gat ggt caa agt gtg gtt tca gta ata gaa cga gt	a 5549

Glu Arg Val Leu Asp Gly Gln Ser Val Val Ser Val Ile Glu Arg Val 1450 1455 1460	
ttt aat aaa ttc cat tta atg acc aaa gaa gca aaa aca caa gat aat Phe Asn Lys Phe His Leu Met Thr Lys Glu Ala Lys Thr Gln Asp Asn 1465 1470 1475	5597
aag aga att aca aca gaa att act tac aat gag gat cta tca aaa agt Lys Arg Ile Thr Thr Glu Ile Thr Tyr Asn Glu Asp Leu Ser Lys Ser 1480 1485 1490 1495	5645
ttc tca gag caa cca gaa aat tta caa caa cct tct cgc gtg tta acc Phe Ser Glu Gln Pro Glu Asn Leu Gln Gln Pro Ser Arg Val Leu Thr 1500 1505 1510	5693
cgt tat acg gat ata caa aca aat act tca cga gaa gag act gtc aat Arg Tyr Thr Asp Ile Gln Thr Asn Thr Ser Arg Glu Glu Thr Val Asn 1515 1520 1525	5741
att aaa agt gat gat tgg gga aat act cta ctt att act gag acc agt Ile Lys Ser Asp Asp Trp Gly Asn Thr Leu Leu Ile Thr Glu Thr Ser 1530 1535 1540	5789
ggg ata cag aaa gaa tac gtt tat tat ccg gtc aat ggc gaa ggt aat Gly Ile Gln Lys Glu Tyr Val Tyr Tyr Pro Val Asn Gly Glu Gly Asn 1545 1550 1555	5837
agt tgc cct gcc gat ccc ttg ggt ttt tct cgg ttc tta aaa tca gtt Ser Cys Pro Ala Asp Pro Leu Gly Phe Ser Arg Phe Leu Lys Ser Val 1560 1565 1570 1575	5885
acg caa aaa gga tcg cct gat gct gct caa agt gtc gca aat aaa gtg Thr Gln Lys Gly Ser Pro Asp Ala Ala Gln Ser Val Ala Asn Lys Val 1580 1585 1590	5933
att cat tat aca tat caa aaa ttt cct act ttt acc ggc gct tat gtt Ile His Tyr Thr Tyr Gln Lys Phe Pro Thr Phe Thr Gly Ala Tyr Val 1595 1600 1605	5981
aag gaa tat gtc agt aaa gtc tca gag acg ata gac aat aaa ata gcg Lys Glu Tyr Val Ser Lys Val Ser Glu Thr Ile Asp Asn Lys Ile Ala 1610 1615 1620	6029
aga acc ttt agc tat gtt aac tca ccg acg agt aaa tct cat ggt tcg Arg Thr Phe Ser Tyr Val Asn Ser Pro Thr Ser Lys Ser His Gly Ser 1625 1630 1635	6077
tta gca aaa ata acg tca gtg atg aat aac cag caa acg gtc acc aca Leu Ala Lys Ile Thr Ser Val Met Asn Asn Gln Gln Thr Val Thr 1640 1645 1650 1655	6125
ttt aaa tat gaa tat tca gaa agt gag atg acc aca aat gct acg gtg Phe Lys Tyr Glu Tyr Ser Glu Ser Glu Met Thr Thr Asn Ala Thr Val 1660 1665 1670	6173
acc ggt ttt gat ggc gca cat atg gaa tcg aaa aat gtg acg tct att Thr Gly Phe Asp Gly Ala His Met Glu Ser Lys Asn Val Thr Ser Ile 1675 1680 1685	6221
tat acc cat cgg caa ctt cgt aaa gtt gat gta aac cac gtg att acc Tyr Thr His Arg Gln Leu Arg Lys Val Asp Val Asn His Val Ile Thr 1690 1695 1700	6269
gat cag tot tat gat out ttg ggt ogo att aca ggg caa att att gat Asp Gln Ser Tyr Asp Leu Leu Gly Arg Ile Thr Gly Gln Ile Ile Asp	6317

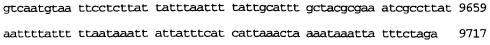
1705			1710		3	L 71 5			
ccc ggc Pro Gly 1720	acg go Thr Al	la Arg	gaa att Glu Ile 725	aaa cgt Lys Arg	aat tac Asn Tyr 1730	gtt tat Val Tyr	caa tat Gln Tyr 1	ccc 6365 Pro 735	
ggc ggt Gly G ly	gac ga 'Asp Gl	aa aat lu Asn . 1740	gat ttt Asp Phe	tgg ccg Trp Pro	gtg atg Val Met 1745	ata gaa Ile Glu	gtt gat Val Asp 1750	tct 6413 Ser	
		g Arg			Asp Gly	Met Gly	cgt att Arg Ile 1765		
			Asp Asp				tcg ggg Ser Gly		
	Gly Th				Ala Arg		gat gtt Asp Val		
		er Lys					aat tta Asn Leu 1		
							acg aaa Thr Lys 1830		
	_	sp Gly			ı Tyr Ser	Thr Glu	tac agt Tyr Ser 1845	-	
							aca att Thr Ile		
	y Val L			Met Le			aat aat Asn Asn		
		la Ser				Asp Gly	acg ata Thr Ile		ř
							gtg acg Val Thr 1910		i
	p Āla G				n Ile Gly		gtg ttt Val Phe 1925		
							a gaa tcc 1 Glu Ser))
	a Ser F			ı Glu Le			g aac gtg 1 Asn Val		7
		eu Gly				/ Leu Gly	g cgg gta y Arg Val		5

agt gat acg gtg ggt cgc aaa acg gaa tat tta tat ggg cct caa Ser Asp Thr Val Gly Gly Arg Lys Thr Glu Tyr Leu Tyr Gly Pro Gln 1980 1985 1990	7133
ggt gac aaa ccg att cag tca att act cct tcg cat aat aag caa aat Gly Asp Lys Pro Ile Gln Ser Ile Thr Pro Ser His Asn Lys Gln Asn 1995 2000 2005	7181
atg gat tac ctc tac tat ctt ggt agt gtg atg tcc aaa ttt acc acg Met Asp Tyr Leu Tyr Tyr Leu Gly Ser Val Met Ser Lys Phe Thr Thr 2010 2015 2020	7229
ggg aca gac caa caa aac ttt cgt tat cat tcg aaa acg gga aca tta Gly Thr Asp Gln Gln Asn Phe Arg Tyr His Ser Lys Thr Gly Thr Leu 2025 2030 2035	7277
tta tct gcg tca gaa ggc gta tct cag act aat tac agt tat ttc cca Leu Ser Ala Ser Glu Gly Val Ser Gln Thr Asn Tyr Ser Tyr Phe Pro 2040 2045 2050 2055	7325
tcg ggt gta tta cag cga gaa tca ttt tta cgg gat aat aaa ccg att Ser Gly Val Leu Gln Arg Glu Ser Phe Leu Arg Asp Asn Lys Pro Ile 2060 2065 2070	7373
tca tcg ggc gag tac ctt tat acg atg tcc ggt ttg att caa cgt cat Ser Ser Gly Glu Tyr Leu Tyr Thr Met Ser Gly Leu Ile Gln Arg His 2075 2080 2085	7421
aaa gat agt ttt ggt cat aat cat gtt tat agt tac gat gct cag gga Lys Asp Ser Phe Gly His Asn His Val Tyr Ser Tyr Asp Ala Gln Gly 2090 2095 2100	7469
aga ttg gtc aaa aca gaa cag gat gca caa tac gct aca ttt gaa tat Arg Leu Val Lys Thr Glu Gln Asp Ala Gln Tyr Ala Thr Phe Glu Tyr 2105 2110 2115	7517
gac aat gtt ggg cga ttg ata aca acg acg acc aaa gac acg acg tca Asp Asn Val Gly Arg Leu Ile Thr Thr Thr Thr Lys Asp Thr Thr Ser 2120 2125 2130 2135	7565
tta tcc caa tta gtg aca aaa atc gaa tat gat gct ttt gat cga gaa Leu Ser Gln Leu Val Thr Lys Ile Glu Tyr Asp Ala Phe Asp Arg Glu 2140 2145 2150	7613
ata aaa cgc tcg cta att agt gac ttc tca ata caa gtt att acc tta Ile Lys Arg Ser Leu Ile Ser Asp Phe Ser Ile Gln Val Ile Thr Leu 2155 2160 2165	7661
agc tat acg aag aat aat caa atc agt caa cgt atc acc tcc atc gat Ser Tyr Thr Lys Asn Asn Gln Ile Ser Gln Arg Ile Thr Ser Ile Asp 2170 2175 2180	7709
ggg gtg gtt atg aaa aat gaa cgt tat caa tat gat aat aat caa cgc Gly Val Val Met Lys Asn Glu Arg Tyr Gln Tyr Asp Asn Asn Gln Arg 2185 2190 2195	7757
tta agc caa tac caa tgt gag gga gaa caa tct ccg att gat cat acg Leu Ser Gln Tyr Gln Cys Glu Gly Glu Gln Ser Pro Ile Asp His Thr 2200 2205 2210 2215	7805
ggt cgt gta tta aat cag cag att tac cat tat gac caa tgg gga aat Gly Arg Val Leu Asn Gln Gln Ile Tyr His Tyr Asp Gln Trp Gly Asn 2220 2235 2230	7853

att aag cgg ctc gat aat aca tat cga gat ggt aag gaa acg gtg gat Ile Lys Arg Leu Asp Asn Thr Tyr Arg Asp Gly Lys Glu Thr Val Asp 2235 2240 2245	7901
tat cat ttc agt caa gcc gat cca act caa ctt att cgt att acc agc Tyr His Phe Ser Gln Ala Asp Pro Thr Gln Leu Ile Arg Ile Thr Ser 2250 2255 2260	7949
gac aaa cag cag ata gag tta agt tat gat gct aat ggc aac cta aca Asp Lys Gln Gln Ile Glu Leu Ser Tyr Asp Ala Asn Gly Asn Leu Thr 2265 2270 2275	7997
cgt gac gaa aaa ggg caa acg ctc att tac gat cag aat aat cgc ttg Arg Asp Glu Lys Gly Gln Thr Leu Ile Tyr Asp Gln Asn Asn Arg Leu 2280 2285 2290 2295	8045
gta cag gtc aaa gac cgg ttg ggc aat ctg gtg tgc agc tac cag tat Val Gln Val Lys Asp Arg Leu Gly Asn Leu Val Cys Ser Tyr Gln Tyr 2300 2305 2310	8093
gat gca ttg aac aaa tta acc gca cag gtt ttg gcg aat ggt acc gtt Asp Ala Leu Asn Lys Leu Thr Ala Gln Val Leu Ala Asn Gly Thr Val 2315 2320 2325	8141
aat cga cag cat tat gct tcc ggt aaa gtg acg aat att caa ttg ggt Asn Arg Gln His Tyr Ala Ser Gly Lys Val Thr Asn Ile Gln Leu Gly 2330 2335 2340	8189
gat gaa gcg att act tgg ttg agc agt gat aag caa cga att gga cat Asp Glu Ala Ile Thr Trp Leu Ser Ser Asp Lys Gln Arg Ile Gly His 2345 2350 2355	8237
caa agc gcc aag aat ggt caa tca gtc tac tat caa tat ggt att gac Gln Ser Ala Lys Asn Gly Gln Ser Val Tyr Tyr Gln Tyr Gly Ile Asp 2360 2365 2370 2375	8285
cat aac agt acg gtt atc gcc agt cag aac gaa aac gag ttg atg gct His Asn Ser Thr Val Ile Ala Ser Gln Asn Glu Asn Glu Leu Met Ala 2380 2385 2390	8333
tta tcc tat aca cct tat ggc ttt agg agt tta att tcc tca tta ccg Leu Ser Tyr Thr Pro Tyr Gly Phe Arg Ser Leu Ile Ser Ser Leu Pro 2395 2400 2405	8381
ggt ttg aat ggc gca cag gtt gat cca gta aca ggc tgg tac ttc tta Gly Leu Asn Gly Ala Gln Val Asp Pro Val Thr Gly Trp Tyr Phe Leu 2410 2415 2420	8429
ggt aac gga tat cgt gtt ttc aac ccg gtt ctc atg agg ttt cac agc Gly Asn Gly Tyr Arg Val Phe Asn Pro Val Leu Met Arg Phe His Ser 2425 2430 2435	8477
ccc gat agt tgg agt cct ttt ggt cgg gga ggg att aac cct tat accPro Asp Ser Trp Ser Pro Phe Gly Arg Gly Gly Ile Asn Pro Tyr Thr2440244524502455	8525
tat tgc caa ggc gat ccc ata aac cgg att gat ctg aac ggt cat ctt Tyr Cys Gln Gly Asp Pro Ile Asn Arg Ile Asp Leu Asn Gly His Leu 2460 2465 2470	8573
agt gcc ggc ggg ata tta ggc att gtg cta ggg gca att ggc atc att Ser Ala Gly Gly Ile Leu Gly Ile Val Leu Gly Ala Ile Gly Ile Ile 2475 2480 2485	8621
gtc ggg att gta tca ctg gga gcc gga gcg gcg att agc gcg ggt ctc	8669



Val Gly Ile Val Ser Leu Gly Ala Gly Ala Ala Ile Ser Ala Gly Leu 2490 2495 2500
att get geg ggg gge get ttg ggg geg att get tet acc age geg ett Ile Ala Ala Gly Gly Ala Leu Gly Ala Ile Ala Ser Thr Ser Ala Leu 2505 2510 2515
gca gtt act gcg act gtc att gga ttg gct gcc gat tcg ata ggg att Ala Val Thr Ala Thr Val Ile Gly Leu Ala Ala Asp Ser Ile Gly Ile 2520 2535 2530 2535
gcg tca gca gca tta tcg gaa aaa gat ccg aaa aca tct ggg ata tta 8813 Ala Ser Ala Ala Leu Ser Glu Lys Asp Pro Lys Thr Ser Gly Ile Leu 2540 2545 2550
aat tgg att agt gcg gga ttg ggg gtt tta agc ttt ggt atc agc gca 8861 Asn Trp Ile Ser Ala Gly Leu Gly Val Leu Ser Phe Gly Ile Ser Ala 2555 2560 2565
ata acc ttt acc tct tcg ctg gta aaa tcg gca cgg agt ggt tct cag 8909 Ile Thr Phe Thr Ser Ser Leu Val Lys Ser Ala Arg Ser Gly Ser Gln 2570 2575 2580
gca gtc agc gcg ggt gtt atc ggg tca gtg cct ctt gaa ttt ggt gaa 8957 Ala Val Ser Ala Gly Val Ile Gly Ser Val Pro Leu Glu Phe Gly Glu 2585 2590 2595
gtt gct agc cgt tcc agc aga cga tgg gat att gcg tta tct tcg ata 9005 Val Ala Ser Arg Ser Ser Arg Arg Trp Asp Ile Ala Leu Ser Ser Ile 2600 2605 2610 2615
teg ttg ggc gca aat gcg gcg tet ete tet acg ggg ata gcg gcg gcg 9053 Ser Leu Gly Ala Asn Ala Ala Ser Leu Ser Thr Gly Ile Ala Ala Ala 2620 2625 2630
gcg gtt gca gac agt aat gcg aat gca gct aat att ctg gga tgg gta 9101 Ala Val Ala Asp Ser Asn Ala Asn Ala Asn Ile Leu Gly Trp Val 2635 2640 2645
tcc ttt ggt ttt ggt gca gta tcg aca acc tca gga ata att gag ctt 9149 Ser Phe Gly Phe Gly Ala Val Ser Thr Thr Ser Gly Ile Ile Glu Leu 2650 2655 2660
acg cgt aca gct tat gca gtg aat cat cag act tgg gaa ctg agt tca 9197 Thr Arg Thr Ala Tyr Ala Val Asn His Gln Thr Trp Glu Leu Ser Ser 2665 2670 2675
tca gca ggt act tcg gag gaa gtg aag cct ata cgt tgt ctc gtt tca 9245 Ser Ala Gly Thr Ser Glu Glu Val Lys Pro Ile Arg Cys Leu Val Ser 2680 2685 2690 2695
cac cgc tgg aat cag aag cag tga atgttaaccc tcctcgggca gttgagttaa 9299 His Arg Trp Asn Gln Lys Gln 2700
tcaaacgttt cgaaatagta ccgggaacta tttagccaat cgtccattga aacccgtaat 9359
gtgttgcgac gtcgtttgac aatataaaga ttctgcgaac cgattggtta agtctcacga 9419
aaaataacta ttaggcgaca tttgcgtcgc cttttttaag gaactttatc aggttacatt 9479
tataagaagc tattttgttt tcgacggatg ttggtttctc tgagataaaa aatagaggga 9539
aatgatgtca agggtgataa tggttaattg taaaatatgt gatattattc gcatttatat 9599



<212	> 41 > PF	T	habo	łus l	umir	nesce	ens								
<400 Met 1		Arg	Ala	Gln 5	Arg	Val	Val	Ile	Thr 10	Gly	Met	Gly	Ala	Val 15	Thr
Pro	Ile	Gly	Glu 20	Asp	Val	Glu	Ser	Cys 25	Trp	Gln	Ser	Ile	Ile 30	Glu	Lys
Gln	His	Arg 35	Phe	His	Arg	Ile	Glu 40	Phe	Pro	Asp	Ser	Phe 45	Ile	Asn	Ser
Arg	Phe 50	Phe	Ser	Phe	Leu	Ala 55	Pro	Asn	Pro	Ser	Arg 60	Tyr	Gln	Leu	Leu
Pro 65	Lys	Lys	Leu	Thr	His 70	Thr	Leu	Ser	Asp	Cys 75	Gly	Lys	Ala	Ala	Leu 80
Lys	Ala	Thr	Tyr	Gln 85	Ala	Phe	Thr	Gln	Ala 90	Phe	Gly	Val	Asn	Ile 95	Ser
Pro	Val	Glu	Tyr 100	Tyr	Asp	Lys	Tyr	Glu 105	Cys	Gly	Val	Ile	Leu 110	Gly	Ser
Gly	Trp	Gly 115	Ala	Ile	Asp	Asn	Ala 120	Gly	Asp	His	Ala	Cys 125	Gln	Tyr	Lys
Gln	Ala 130	Lys	Leu	Ala	His	Pro 135	Met	Ser	Asn	Leu	Ile 140	Thr	Met	Pro	Ser
Ser 145	Met	Thr	Ala	Ala	Cys 150	Ser	Ile	Met	Tyr	Gly 155	Leu	Arg	Gly	Tyr	Glr 160
Asn	Thr	Val	Met	Ala 165	Ala	Cys	Ala	Thr	Gly 170	Thr	Met	Ala	Ile	Gly 175	Asp
Ala	Phe	Glu	Ile 180	Ile	Arg	Ser	Gly	Arg 185	Ala	Lys	Cys	Met	Ile 190	Ala	Gly
Ala	Ala	Glu 195	Ser	Leu	Thr	Arg	Glu 200	Cys	Asn	Ile	Trp	Ser 205	Ile	Asp	Val
Leu	Asn 210	Ala	Leu	Ser	Lys	Glu 215	Gln	Ala	Asp	Pro	Asn 220	Leu	Ala	Cys	Cys
Pro 225	Phe	Ser	Leu	Asp	Arg 230	Ser	Gly	Phe	Val	Leu 235	Ala	Glu	Gly	Ala	Ala 240
Val	Val	Cys	Leu	Glu 245	Asn	Tyr	Asp	Ser	Ala 250	Ile	Ala	Arg	Gly	Ala 255	Thr
Ile	Leu	Ala	Glu 260	Ile	Lys	Gly	Tyr	Ala 265	Gln	Tyr	Ser	Asp	Ala 270	Val	Asr

Leu Thr Arg Pro Thr Glu Asp Ile Glu Pro Lys Ile Leu Ala Ile Thr 275 280 285

Lys Ala Ile Glu Gln Ala Gln Ile Ser Pro Lys Asp Ile Asp Tyr Ile 295 Asn Ala His Gly Thr Ser Thr Pro Leu Asn Asp Leu Tyr Glu Thr Gln 310 Ala Ile Lys Ala Ala Leu Gly Gln Tyr Ala Tyr Gln Val Pro Ile Ser Ser Thr Lys Ser Tyr Thr Gly His Leu Ile Ala Ala Ala Gly Ser Phe Glu Thr Ile Val Cys Val Lys Ala Leu Ala Glu Asn Cys Leu Pro Ala 360 Thr Leu Asn Leu His Arg Ala Asp Pro Asp Cys Asp Leu Asn Tyr Leu Pro Asn Gln His Cys Tyr Thr Ala Gln Pro Glu Val Thr Leu Asn Ile 395 Ser Ala Gly Phe Gly Gly His Asn Ala Ala Leu Val Ile Ala Lys Val Arg <210> 3 <211> 253 <212> PRT <213> Photorhabdus luminescens <400> 3 Met Glu Asp Ile Glu His Trp Ser Asn Phe Ser Gly Asp Phe Asn Pro Ile His Tyr Ser Ala Lys Ser Glu Ser Leu Arg Asn Ile Gln Gln His Pro Val Gln Gly Met Leu Ser Leu Leu Tyr Val Arg Gln Gln Phe Ser Gln Leu Thr Ser Ala Phe Thr Thr Gly Ile Leu Asn Ile Asp Ala Ser Phe Arg Gln Tyr Val Tyr Thr Ala Leu Pro His Gln Leu Arg Ile Asn Thr Lys Asn Lys Thr Phe Lys Leu Glu Asn Pro Ser Lys Glu Asn Thr Leu Phe Gly Asn Thr Ser Val Glu Asn Thr Met Glu Ser Ile Glu Asp 105 Trp Ile Val Gln Asp Asn Cys Gln Lys Leu Thr Ile Thr Gly Glu Glu 120 Val Cys Glu Lys Tyr Ala Val Phe Arg Tyr Tyr Phe Pro Ser Val Thr

Ser Ile Gly Trp Phe Leu Asp Ala Leu Ala Phe His Leu Ile Ile Asn

Ser Thr Gly Phe Leu Asn Phe Glu His Tyr His Phe Asn Gln Leu Gln

155

170

Asp Tyr Leu Ser Gln Ser Phe Thr Leu His Thr Gly Gln Ala Ile Lys 180 185 190

Ile Arg Lys Glu Ile Val Asn Ser Thr Val Leu Leu Ser Ser Pro Asp

195 200 205

Ile Cys Val Glu Leu Asn Pro Pro Leu Leu Ile Lys Asn Gly Asp Lys 210 215 220

Asp Tyr Ile Arg Ile Phe Tyr Tyr Arg Cys Leu Tyr Asp Lys Lys Pro 225 230 235 240

Ile Phe Val Ser Lys Thr Ser Ile Ile Ser Lys Met Lys 245 250

<210> 4

<211> 186

<212> PRT

<213> Photorhabdus luminescens

<400> 4

Met Asn Val Leu Glu Gln Gly Lys Val Ala Ala Leu Tyr Ser Ala Tyr 1 5 10 15

Ser Glu Thr Glu Gly Ser Ser Trp Val Gly Asn Leu Cys Cys Phe Ser 20 25 30

Ser Asp Arg Glu His Leu Pro Ile Ile Val Asm Gly Arg Arg Phe Leu 35 40 45

Ile Glu Phe Val Ile Pro Asp His Leu Leu Asp Lys Thr Val Lys Pro 50 60

Arg Val Phe Asp Leu Asp Ile Asn Lys Gln Phe Leu Leu Arg Arg Asp 65 70 75 80

His Arg Glu Ile Asn Ile Tyr Leu Leu Gly Glu Gly Asn Phe Met Asp 85 90 95

Arg Thr Thr Asp Lys Asn Leu Phe Glu Leu Asn Glu Asp Gly Ser 100 105 110

Leu Phe Ile Lys Thr Leu Arg His Ala Leu Gly Lys Tyr Val Ala Ile 115 120 125

Asn Pro Ser Thr Thr Gln Phe Ile Phe Phe Ala Gln Gly Lys Tyr Ser 130 140

Glu Phe Ile Met Asn Ala Leu Lys Thr Val Glu Asp Glu Leu Ser Lys 145 150 155 160

Arg Tyr Arg Val Arg Ile Ile Pro Glu Leu Gln Gly Pro Tyr Tyr Gly 165 170 175

Phe Glu Leu Asp Ile Leu Ser Ile Thr Ala 180 185

<210> 5

<211> 258

<212> PRT

<213> Photorhabdus luminescens

- 16 -

<400> 5 Met Glu Lys Lys Ile Thr Thr Phe Thr Ile Glu Lys Thr Asp Asp Asn Phe Tyr Ala Asn Gly Arg His Gln Cys Met Val Lys Ile Ser Val Leu Lys Gln Glu Tyr Arg Asn Gly Asp Trp Ile Lys Leu Ala Leu Ser Glu Ala Glu Lys Arg Ser Ile Gln Val Ala Ala Leu Ser Asp Ser Leu Ile Tyr Asp Gln Leu Lys Met Pro Ser Gly Trp Thr Thr Thr Asp Ala Arg Asn Lys Phe Asp Leu Gly Leu Leu Asn Gly Val Tyr His Ala Asp Ala Phe Ile Asp Glu Gln Val Thr Asp Arg Ala Gly Asp Cys Cys 100 Thr Asn Glu Asn Tyr Gln Asn Ser Val Lys Ser Val Pro Glu Ile Ile Tyr Arg Tyr Val Ser Ser Asn Arg Thr Ser Thr Glu Tyr Leu Met Ala Lys Met Thr Phe Glu Asp Thr Asp Gly Lys Arg Thr Leu Thr Thr Asn Met Ser Val Gly Asp Glu Val Phe Asp Ser Lys Val Leu Leu Lys Ala Ile Ala Pro Tyr Ala Ile Asn Thr Asn Gln Leu His Glu Asn Ile Asn Thr Leu Phe Asp Lys Thr Glu Glu Pro Thr Lys Ser Asp Thr His His 200 Gln Ile Ile Asn Leu Tyr Arg Trp Thr Leu Pro Tyr His Leu Arg Ile 215 Leu Glu Gly Asn Asp Ser Thr Val Asn Arg Ile Tyr Val Leu Gly Lys 230 Glu Pro Ser Asn Asp Arg Phe Leu Thr Arg Gly Arg Val Phe Lys Arg Gly Thr His Met <210> 6 <211> 1584 <212> PRT <213> Photorhabdus luminescens

Met Lys Ala Thr Asp Ile Tyr Ser Asn Ala Phe Asn Phe Gly Ser Tyr

Ile Asn Thr Gly Val Asp Pro Arg Thr Gly Gln Tyr Ser Ala Asn Ile

SDOCID: <WO___9942589A2_I_>

<400> 6

Asn	Ile	Ile 35	Thr	Leu	Arg	Pro	Asn 40	Asn	Val	Gly	Asn	<i>S</i> er 45	Glu	Gln	Thr
Leu	Ser 50	Leu	Ser	Phe	Ser	Pro 55	Leu	Thr	Thr	Leu	Asn 60	Asn	Gly	Phe	Gly
Ile 65	Gly	Trp	Arg	Phe	Ser 70	Leu	Thr	Thr	Leu	Asp 75	Ile	Lys	Thr	Leu	Thr 80
Phe	Ser	Arg	Ala	Asn 85	Gly	Glu	Gln	Phe	Lys 90	Cys	Lys	Pro	Leu	Pro 95	Pro
Asn	Asn	Asn	Asp 100	Leu	Ser	Phe	Lys	Asp 105	Lys	Lys	Leu	Lys	Asp 110	Leu	Arg
Val	Tyr	Lys 115	Leu	Asp	Ser	Asn	Thr 120	Phe	Tyr	Val	Tyr	Asn 125	Lys	Asn	Gly
Ile	Ile 130	Glu	Ile	Leu	Lys	Arg 135	Ile	Gly	Ser	Ser	Asp 140	Ile	Ala	Lys	Thr
Val 145	Ala	Leu	Glu	Phe	Pro 150	Asp	Gly	Glu	Ala	Phe 155	Asp	Leu	Ile	Tyr	Asn 160
Ser	Arg	Phe	Ala	Leu 165	Ser	Glu	Ile	Lys	Тут 170	Arg	Val	Thr	Gly	Lys 175	Thr
Tyr	Leu	Lys	Leu 180	Asn	Tyr	Ser	Gly	Asn 185	Asn	Cys	Thr	Ser	Val 190	Glu	Tyr
Pro	Asp	Asp 195		Asn	Ile	Ser	Ala 200		Ile	Ala	Phe	Asp 205	_	Arg	Asn
Asp	Tyr 210	Leu	Ile	Thr	Val	Thr 215		Pro	Tyr	Asp	Ala 220	Ser	Gly	Pro	Ile
Asp 225		Ala	Arg	Phe	Lys 230	Met	Thr	Tyr	Gln	Thx 235		Lys	Gly	Val	Phe 240
Pro	Val	Ile	Ser	Thr 245	Phe	Arg	Thr	Pro	Thr 250		Tyr	Val	Glu	Leu 255	
Ser	Tyr	Lys	Glu 260		Gly	His	Lys	265		Asp	Thr	Glu	Тут 270		Pro
Tyr	Ala	Ala 275		Leu	Thr	Il∈	280		Gly	/ Asr	Gly	Glr. 285		Ala	Val
Ser	Lys 290		Tyr	Glu	Tyr	Ser 295		· Val	. His	s Asr	Phe 300		ı Gly	Tyr	· Ser
Ser 305	_	Arg	; Thr	Ser	? Phe 310	-	Ser	: Sex	Glr	Asp 315		Lev	ı Tyr	Leu	Val 320
Thr	Gly	Lys	Tyr	Thr 325	Tyr	Ser	Ser	: Ile	330 330		y Val	. Leu	ı Asp	335	
Ser	· Val	. Val	Ser 340		. Ile	: Glu	ı Arg	y Val 345		e Asr	Lys	s Phe	His 350		Met
Thr	: Lys	Glu 355		i Lys	Thr	Glr	360		ı Lys	s Arg	j Ile	365		Glu	ı Ile
Thu	Tyr 370		n Glu	ı Asp	Leu	375		s Sei	r Phe	e Sei	Glu 380		n Pro	Glu	ı Asr

Leu Gln Gln Pro Ser Arg Val Leu Thr Arg Tyr Thr Asp Ile Gln Thr Asn Thr Ser Arg Glu Glu Thr Val Asn Ile Lys Ser Asp Asp Trp Gly Asn Thr Leu Leu Ile Thr Glu Thr Ser Gly Ile Gln Lys Glu Tyr Val Tyr Tyr Pro Val Asn Gly Glu Gly Asn Ser Cys Pro Ala Asp Pro Leu Gly Phe Ser Arg Phe Leu Lys Ser Val Thr Gln Lys Gly Ser Pro Asp Ala Ala Gln Ser Val Ala Asn Lys Val Ile His Tyr Thr Tyr Gln Lys Phe Pro Thr Phe Thr Gly Ala Tyr Val Lys Glu Tyr Val Ser Lys Val Ser Glu Thr Ile Asp Asn Lys Ile Ala Arg Thr Phe Ser Tyr Val Asn 505 Ser Pro Thr Ser Lys Ser His Gly Ser Leu Ala Lys Ile Thr Ser Val Met Asn Asn Gln Gln Thr Val Thr Thr Phe Lys Tyr Glu Tyr Ser Glu Ser Glu Met Thr Thr Asn Ala Thr Val Thr Gly Phe Asp Gly Ala His 550 Met Glu Ser Lys Asn Val Thr Ser Ile Tyr Thr His Arg Gln Leu Arg 565 570 575 Lys Val Asp Val Asn His Val Ile Thr Asp Gln Ser Tyr Asp Leu Leu Gly Arg Ile Thr Gly Gln Ile Ile Asp Pro Gly Thr Ala Arg Glu Ile Lys Arg Asn Tyr Val Tyr Gln Tyr Pro Gly Gly Asp Glu Asn Asp Phe 615 Trp Pro Val Met Ile Glu Val Asp Ser Gln Gly Val Arg Arg Lys Thr His Tyr Asp Gly Met Gly Arg Ile Cys Ser Ile Glu Glu Gln Asp Asp Asp Gly Ala Trp Gly Thr Ser Gly Ile Tyr Gln Gly Thr Tyr Arg Lys Val Leu Ala Arg Gln Tyr Asp Val Leu Gly Gln Leu Ser Lys Glu Ile Ser Asn Asp Trp Leu Trp Asn Leu Ser Ala Asn Pro Leu Val Arg Leu 695 Ala Thr Pro Leu Val Thr Thr Lys Thr Tyr Lys Tyr Asp Gly Trp Gly Asn Leu Tyr Ser Thr Glu Tyr Ser Asp Gly Arg Ile Glu Leu Glu Ile

				725					730					735	
His	Asp	Pro	Ile 740	Thr	Arg	Thr	Ile	Thr 745	Gln	Gly	Val	Lys	Gly 750	Leu	Gly
Met	Leu	Asn 755	Ile	Gln	Gln	Asn	Asn 760	Phe	Glu	Gln	Pro	Ala 765	Ser	Ile	Lys
Ala	Val 770	Tyr	Pro	Asp	Gly	Thr 775	Ile	Tyr	Ser	Thr	Arg 780	Thr	Tyr	Arg	Tyr
Asp 785	Gly	Phe	Gly	Arg	Thr 790	Val	Thr	Glu	Thr	Asp 795	Ala	Glu	Gly	His	Ala 800
Thr	Gln	Ile	Gly	Tyr 805	Asp	Val	Phe	Asp	Arg 810	Ile	Val	Lys	Lys	Thr 815	Leu
Pro	Asp	Gly	Thr 820	Ile	Leu	Glu	Ser	Ala 825	Tyr	Ala	Ser	Phe	Ser 830	His	Glu
Glu	Leu	Ile 835	Ser	Ala	Leu	Asn	Val 840	Asn	Gly	Thr	Gln	Leu 845	Gly	Ala	Leu
Val	Tyr 850	Asp	Gly	Leu	Gly	Arg 855	Val	Ile	Ser	Asp	Thr 860	Val	Gly	Gly	Arg
Lys 865	Thr	Glu	Tyr	Leu	Tyr 870	Gly	Pro	Gln	Gly	Asp 875	Lys	Pro	Ile	Gln	Ser 880
Ile	Thr	Pro	Ser	His 885	Asn	Lys	Gln	Asn	Met 890	Asp	Tyr	Leu	Tyr	Tyr 895	Leu
Gly	Ser	Val	Met 900	Ser	Lys	Phe	Thr	Thr 905	Gly	Thr	Asp	Gln	Gln 910	Asn	Phe
		915				Gly	920					925			
	930					Туг 935					940				
945					950	Lys				955					960
Thr	Met	Ser	Gly	Leu 965	Ile	Gln	Arg	His	Lys 970	Asp	Ser	Phe	Gly	His 975	Asn
			980			Ala		985					990		
Asp	Ala	Gln 995	Tyr	Ala	Thr	Phe	Glu 1000	Tyr	Asp	Asn		Gly 1005	Arg	Leu	Ile
Thr	Thr 1010	Thr	Thr	Lys		Thr 1015	Thr	Ser	Leu		Gln 1020	Leu	Val	Thr	Lys
Ile 025	Glu	Tyr	Asp		Phe 1030	Asp	Arg	Glu		Lys 1035	Arg	Ser	Leu		Ser 1040
Asp	Phe	Ser		Gln 1045	Val	Ile	Thr		Ser 1050		Thr	Lys		Asn 1055	Gln
Ile	Ser		Arg 1060		Thr	Ser		Asp 1065	Gly	Val	Val		Lys 1070	Asn	Glu

- Arg Tyr Gln Tyr Asp Asn Asn Gln Arg Leu Ser Gln Tyr Gln Cys Glu 1075 1080 1085
- Gly Glu Gln Ser Pro Ile Asp His Thr Gly Arg Val Leu Asn Gln Gln 1090 1095 1100
- Ile Tyr His Tyr Asp Gln Trp Gly Asn Ile Lys Arg Leu Asp Asn Thr 105 1110 1115 1120
- Tyr Arg Asp Gly Lys Glu Thr Val Asp Tyr His Phe Ser Gln Ala Asp 1125 1130 1135
- Pro Thr Gln Leu Ile Arg Ile Thr Ser Asp Lys Gln Gln Ile Glu Leu 1140 1145 1150
- Ser Tyr Asp Ala Asn Gly Asn Leu Thr Arg Asp Glu Lys Gly Gln Thr 1155 1160 1165
- Leu Ile Tyr Asp Gln Asn Asn Arg Leu Val Gln Val Lys Asp Arg Leu 1170 1180
- Gly Asn Leu Val Cys Ser Tyr Gln Tyr Asp Ala Leu Asn Lys Leu Thr 185 1190 1195 1200
- Ala Gln Val Leu Ala Asn Gly Thr Val Asn Arg Gln His Tyr Ala Ser 1205 1210 1215
- Gly Lys Val Thr Asn Ile Gln Leu Gly Asp Glu Ala Ile Thr Trp Leu 1220 1225 1230
- Ser Ser Asp Lys Gln Arg Ile Gly His Gln Ser Ala Lys Asn Gly Gln 1235 1240 1245
- Ser Val Tyr Tyr Gln Tyr Gly Ile Asp His Asn Ser Thr Val Ile Ala 1250 1255 1260
- Ser Gln Asn Glu Asn Glu Leu Met Ala Leu Ser Tyr Thr Pro Tyr Gly 265 1270 1275 1280
- Phe Arg Ser Leu Ile Ser Ser Leu Pro Gly Leu Asn Gly Ala Gln Val 1285 1290 1295
- Asp Pro Val Thr Gly Trp Tyr Phe Leu Gly Asn Gly Tyr Arg Val Phe 1300 1305 1310
- Asn Pro Val Leu Met Arg Phe His Ser Pro Asp Ser Trp Ser Pro Phe 1315 1320 1325
- Gly Arg Gly Gly Ile Asn Pro Tyr Thr Tyr Cys Gln Gly Asp Pro Ile 1330 1335 1340
- Asn Arg Ile Asp Leu Asn Gly His Leu Ser Ala Gly Gly Ile Leu Gly 345 1350 1355 1360
- Ile Val Leu Gly Ala Ile Gly Ile Ile Val Gly Ile Val Ser Leu Gly 1365 1370 1375
- Ala Gly Ala Ala Ile Ser Ala Gly Leu Ile Ala Ala Gly Gly Ala Leu 1380 1385 1390
- Gly Ala Ile Ala Ser Thr Ser Ala Leu Ala Val Thr Ala Thr Val Ile 1395 1400 1405
- Gly Leu Ala Ala Asp Ser Ile Gly Ile Ala Ser Ala Ala Leu Ser Glu 1410 1415 1420

- 21 -

Lys Asp Pro Ly 425	s Thr Ser 1430	Gly Ile		Trp Ile 1435	Ser Ala	Gly Leu 1440			
Gly Val Leu Se	er Phe Gly 1445	Ile Ser	Ala Ile 1450	Thr Phe		Ser Leu 1455			
Val Lys Ser Al 146			Gln Ala 1465	Val Ser	Ala Gly 1470	Val Ile			
Gly Ser Val Pr 1475	co Leu Glu	Phe Gly 1480	Glu Val		Arg Ser 1485	Ser Arg			
Arg Trp Asp Il 1490		Ser Ser 1495	Ile Ser	Leu Gly 1500		Ala Ala			
Ser Leu Ser Tr 505	r Gly Ile 1510	Ala Ala		Val Ala 1515	Asp Ser	Asn Ala 1520			
Asn Ala Ala As	an Ile Leu 1525	Gly Trp	Val Ser 1530	Phe Gly		Ala Val 1535			
Ser Thr Thr Se			Leu Thr 1545	Arg Thr	Ala Tyr 1550	Ala Val			
Asn His Gln Th 1555	nr Trp Glu	Leu Ser 1560	Ser Ser		Thr Ser 1565	Glu Glu			
Val Lys Pro II 1570	le Arg Cys	Leu Val 1575	Ser His	Arg Trp 1580		Lys Gln			
<210> 7 <211> 18 <212> DNA <213> Artificial Sequence <220> <223> Description of Artificial Sequence:oligonucleotide									
<400> 7 acacagcagg tto	cgtcag						18		
<210> 8 <211> 18 <212> DNA <213> Artific:	ial Sequen	ce							
<220> <223> Descript	tion of Ar	tificial	Sequenc	e:oligon	ucleotid	е			
<400> 8 ggcagaagca cto	caactc						18		
<210> 9 <211> 20 <212> DNA <213> Artific	ial Sequer	ıce							
<220> <223> Descript	tion of Ar	tificial	Sequenc	e:oligor	ucleotid	e			
<400> 9									

- 22 -

```
20
attgatagca cgcggcgacc
<210> 10
<211> 22
<212> DNA
<213> Artificial Sequence
<220>
<223> Description of Artificial Sequence:oligonucleotide
<400> 10
                                                                   22
ttgtaacgtg gagccgaact gg
<210> 11
<211> 37948
<212> DINA
<213> Photorhabdus luminescens
<220>
<221> CDS
<222> (15171)..(18035)
<223> orf5
<220>
<221> CDS
<222> (23768)..(31336)
<223> hph2
<220>
<221> CDS
<222> (31393)..(35838)
<223> orf2
<400> 11
tgttgctgga ccgtggagat tatgcctatc gtcagttaga acgagacacg ctcaatgaag 60
ccaagatgtg gtatatgcaa gcactgcatc tgttaggcga taaacctcat ctatcgttca 120
 qttcaqaqtq gagcaaaccg agtttaggcg acgctgccgg aacggaaaga caaaagcaac 180
 acgcccaagc aatggccgct ctgcgacaag gtgatgttag tcggcacaac aacccgacag 240
 atcttttctt gccacaggtc aatgaagtga tgcaaaacta ttggcaaaaa ttggaacaac 300
 ggetgtataa cetgegteat aaceteacta ttgaeggeea aeegetaeat etgeetattt 360
 acqctacacc ggcagatcca aaagcattac ttagegcege tgtegetage teegaaggtg 420
 gggtagetet eteacageea titatgteae tgtggegtit eecacacatg etggaaaaeg 480
 cgcgtggtat ggtcagtcag ctcactcaat tcggctctac gctacaaaat attatcgaac 540
 gtcaggatgc ggaagcttta aacacgctct tgcagaatca agcagcagaa ctgatattga 600
 ctcatctcag catacaggac aaaaccatcg cagagctgga tgcggaaaaa atcgtactgg 660
 aaaaatccaa agccggggcg caatcacgct ttgacagcta caaaaagtta tacgacgaaa 720
 atatcaatgc gggtgaaaac cgggctatag cattgcatgc ctccgttgct ggcctcagca 780
 ctgccctgca agcatcacgt ctggcgggcg ctgcgcttga tctggcgccc aacattttcg 840
 gtctcgctga tggcggtagc cgttggggag cgattgccga agcgacaggt aatgttatgg 900
```

aattctccgc	cagtgtgatg	aacaccgaag	cggataaaat	cagccagtca	gaagcctatc	960
gccgtcgccg	tcaggaatgg	gaaatccagc	gtaatcatgc	cgaagcagag	ataaaacaga	1020
tcgatgctca	acttcaatca	ctggcagtac	gccgtgaagc	cgcggtattg	cagaaagcca	1080
gcctaaaaac	ccaacaggaa	cagactcatg	ctcaattgac	tttcctgcaa	cgtaaattca	1140
gtaatcaagc	gttgtactac	tggctacgcg	gtcggctagc	tgctatttac	ttccaatttt	1200
acgatttggc	cgtagcgcgt	tgtctgatgg	ctgaaatggc	ttatcgttgg	gagactaatg	1260
agaccgcggc	aagctttatc	aaacccggcg	cctggcaggg	aacccatgcg	ggtttactgg	1320
ctggtgaaac	cttgatgctg	aatctggcgc	aaatggaaga	tgcccatttg	aggtgggatc	1380
aacgcgctct	ggaagtggaa	cggaccattt	cattgacgca	acactatgga	gcactgccag	1440
aaaaatcgtt	taatttagcc	acacggattt	ctaccctgct	agcaggtggt	acaactgact	1500
ccattgatga	tcatcccgtt	acattagaaa	acgaccaact	tagtgccaaa	atctctctgt	1560
caggtctgtc	attagataat	gactacccag	atggcaacgg	cgtaggcaac	attcgacgca	1620
ttaaacaaat	cagtgtcacc	ttgccagccc	tgttaggacc	atatcaggat	gtacaagcta	1680
ttctgtccta	cggaggaagt	gaaatcggat	tagctgaaag	ctgtaaatca	ctggcgatct	1740
ctcatgggat	caatgacagt	ggtcaattcc	agttggattt	taacaatggt	aagttcctgc	1800
cgtttgaagg	gattgcgatt	aacgatactg	gcacattgac	actcagtttc	ccccaatgcg	1860
actgtcaaac	aagaaaacat	gttgcagact	ttgagtgata	ttattctgca	tattcgctat	1920
accatecgcc	aataaccacc	tcaattaaat	ассаааааса	ggctcctaaa	cggggcctga	1980
acttttcacg	aatatatacc	actcacagtc	tgctctcttt	acctgtctga	cgctcgttat	2040
aacagagata	tttccttttc	tcgtgagtcc	catcacctac	tataaaatat	caaccctctt	2100
ctttttcata	atatgcaata	tgtaacaaat	gcaattattt	catttagtta	ttgttaacta	2160
gttatattac	ttatgatgta	attataaatt	ttgttattgc	atcacaatag	ccatttaaat	2220
aaataataac	gttgtgaaat	agttgatagt	taaatggtgt	ttttatttag	ccgttatttt	2280
caacccaatt	tcagaccgct	atcagacgtt	acctgtgttg	cctttgtttt	gatagatata	2340
aataacctta	tttatatcca	cggtactcag	accagcataa	atgttttatt	tacctaacat	2400
ttaaaaggaa	taaacatgaa	cacactcaaa	tccgaatatg	aaaacgcgtt	agtagcaggt	2460
tttaataatc	taaccgatat	ttgtcatctc	tcttttgacg	aacttcgcaa	aaaagtgaag	2520
gacaaactct	catggtcaca	gacccaaagc	ttatatcttg	aagcacagca	agtgcaaaag	2580
gacaatctcc	tgcatgaagc	ccatattctg	aaacgcgcca	atcctcattt	acaaagtgcg	2640
gtccatcttg	ccctgacaac	acctcatgct	gaccagcaag	gttataatag	cagatttggc	2700
aatcgcgcca	gcaaatatac	agccccaggc	gcaatttctt	ccatgttttc	teetgegget	2760
tatttagctg	aactttatcg	tcaggcacgg	aatttacatg	atgaaaattc	tatttatcat	2820

		tataaaataa	ttaataataa	4443433333	tataastaa	2880
	ggcgtccgga					
	cactttctct					
	tgctggaggc					
gaggctttga	agaatattaa	acctctgtcc	aaggacgatc	tgctggaggc	tttgaagaat	3060
attaaacctc	tgtccaagga	cgatctgctg	gaggctttga	agaatattaa	acctctgtcc	3120
aaggacgatc	tgctggaggc	tttgaagaat	attaaacctc	tgtccaagga	cgatctgctg	3180
gaggctttga	agaatattaa	acctctgtcc	aaggacgatc	tgctggaggc	tttgaagaat	3240
attaaacctc	tgtccaagga	cgatctgctg	gaggctttga	agaatattaa	acctctgtcc	3300
aaggacgato	tacaggaatg	tattgaaatt	ctattcaatc	tggacagcca	cactaaagta	3360
atgaaagcgt	tatccaattt	ccgcgtttct	ggcatgatgc	catatcacga	tgcttatgaa	3420
agcgtgcgta	aggttgttca	attacaggct	ccggtgtttg	aacacgttgt	tagtacatca	3480
ctagaaacga	ctatcgatga	actaaaatat	caagcttctt	tgttggaaat	taattcttct	3540
gtctcgccta	aattatttac	tatcttgact	gaagaaatta	ctacaatcaa	tgcaagaagt	3600
ctctatgagg	g aaaattttgg	taatattaaa	ccttctctaa	taggaaaacc	ggaatatctg	3660
aaaagttatt	acaatctgag	tgatgaagag	tttagcgatt	tcattaaaat	aagaactata	3720
cttcttccag	g aagaagaaat	agcaattact	gatettgeat	. cgcgtactac	tagtacacaa	3780
cagactatco	g aaaatcctga	ttatcgtgct	. ctattgaaaa	ttaataagtt	tattcgtcta	3840
ttcaaagct	a taaacttato	accgacggta	ttaagtggaa	tcctccgcag	catcagcaca	3900
gaattcaata	a tcaataaaga	aatattacaa	aaaatctttc	gtgttaaata	ctatatgcaa	3960
cgttatggta	a ttgacactga	gactgcatta	atactatgca	aggtaccaat	ttcacaatat	4020
atcaatgac	g gacatctaag	tcagtttgat	cgtttattta	attccccaa	actgaatggt	4080
caagatttt	t ccgtcaatgg	g tactcagaat	attgatttaa	a ccctaagcag	taccaacaa	4140
tggaataaa	a cagtacttaa	a acgtgctttt	. aacctcgatg	g atatctcatt	. aaatcgacta	a 4200
ctaaaaatt	a ccaatccggt	caatactacc	gaaatgataa	a ctaatgatat	agagaatct	4260
tctcatctc	t ataggacaaa	a attactggca	a gatatccato	e agttaactat	tgatgaact	g 4320
gggttactg	t tggaagccat	t aggtaaagga	a acaaccaatt	tatetgagat	tactcctga	2 4380
aatctggtt	a ctctaattaa	a caaactcta	ge tg teacta	a getggetaeg	g tacacaaaa	g 4440
tggagtgtc	t atcagttgt	t tatgatgaci	t actgataaa	t ataacaaaa	c cctaacccc	g 4500
gaaatacgg	a atttactgg	a taccgtcta	c aatggcttg	c aagattttaa	a taagaagat	g 4 560
ttgaaagct	g aagaagatc	t agagaaaac	c aaaaagaaa	t tgcagagcgo	c caaggaaaa	t 4 620
ctggaaaaa	t teceggaaa	a ccagccaca	a ctccaagaa	g acaggaaaaa	a agcccagag	a 4 680
agactgaat	a aagctgaag	a gacccacga	a aaagccgaga	a aaaacctaga	a tgaggtcag	g 4 740
aaaaatctg	c caaaagcca	t atctcctta	t ategeegee	g ctctgcaat!	t accatctga	a 4800

catgcggcat	attccatact	catctgggca	gataatctgg	aacccggcat	aggaaaaatg	4860
acagcggaaa	aattatggaa	ctggttg c gg	aaaaatcccg	ttacggctca	acctgaattc	4920
caaaaacaag	ctgaacctgt	ggtccagtat	tgccagcgcc	tggcacaact	agcgttgatt	4980
taccgttcta	ccggccttaa	cgaaaacacc	ttaagtctgt	ttgtgacaaa	gccgcaacac	5040
tttgttatta	aaaccaaagc	acccgaaaca	actgaaacaa	caccagcaca	tgacgtatca	5100
acactaatgt	cactaacgcg	ttttactgac	tgggttaact	cactaggtga	aaacgcctct	5160
tctgtactaa	ccgaatttaa	aaaaggaaca	ttaacagcag	aactattggc	taaggctatg	5220
aatcttgata	aaaatctact	ggagcaagcc	aagattcagg	caaaaactga	tttgtccaac	5280
tggccatcta	tcgacaacct	attgcagtgg	attaacatct	cacgtcaatt	gaacatctct	5340
ccacaaggca	tttccacact	gactcaagta	ttgaccgcag	aacctcccgc	taactatacc	5400
caatgggaaa	acgccgctgc	gatattaacc	gccgggctgg	acacccaaaa	gactaacgcc	5460
ctacatgcgt	ttctggatga	gtctcgcagt	gctgcgttaa	gcacatacta	tatttattct	5520
cataaccaaa	aagatcgaga	agcaagaaaa	catacagtaa	ttaaaaaccg	tgatgatcta	5580
tatcaatacc	tattgatcga	taaccaagtt	teegeegaea	ttaaaactac	agagatcgct	5640
gaagctatcg	ctagtatcca	actgtatatt	aaccgcgcgt	tgaaaaatat	ggagggagat	5700
actgtcacaa	gtgtcaccag	ccgctcattc	ttcaccaact	gggataaata	caataaacgc	5760
tacagcactt	gggccggtat	ggctaaactc	ctttactatc	cagagaatta	catcgatccg	5820
acgctacgta	ttgggcagac	aaaaatgatg	gatacgttgc	tgcaatccat	cagccaaagc	5880
caattaaata	tcgataccgt	agaagatgcc	tttaaatctt	acctaacatc	attcgaacag	5940
gtggctaatc	tggaaatcct	cagcgcctac	catgacaaca	ttaataatga	tcaaggatta	6000
acttacttta	tcggacgtag	taaaacagaa	gtgaatcaat	attattggcg	cagtgtggat	6060
cacaataaat	ccagcgaagg	taaattcccc	gctaatgcct	ggagtgagtg	gtacaaaatt	6120
gattgtccaa	ttaaccccta	caaagatact	attcacccgg	taattttcca	atctcgcctg	6180
tatcttatct	ggctggaaca	aaaaaaggcg	actaaacagg	aaggtgataa	aaccgcctcg	6240
ggttattatt	atgaactgaa	attagcgcat	atccgttatg	acggcacctg	gaatacacca	6300
gtcacctttg	atgtaaacca	aaaaatatco	gatttaaatc	tgggaaataa	aacacctgga	6360
ctttactgct	caagctttca	aggcagagat	. gaattgctgg	tgatgtttta	taaaaaacaa	6420
gatcaattaa	atcaatacac	: aaacacagta	ccaataaaag	gactatatat	cacttccaat	6480
atgtcttcta	aggaaatgac	acctgaaaat	cacaaaccta	acgettataa	acagtttgat	6540
actaatagta	ttattggtgt	. caataatcgc	tatgcagaaa	gctacgaaat	cccttcatca	6600
gtaaatagta	ataacggtta	tgattgggga	gatggctatc	tgagtatggt	gtatggcgga	6660
aatatttcag	ccatcaaact	ggagteetea	tcagataagt	. taaaactctc	accaaggtta	6720

agaattattc	ataatggact	tgtaggccga	caacgcaacc	aatgcaacct	gatgaagaaa	6780
tacggtcagc	ttggtgataa	atttattatt	tatactactc	taggtattaa	ccccaataat	6840
ttgtcgaata	aaaaattcat	ttaccctgtt	tatcagtata	gtgggaacac	taccaataat	6900
gagaaaggac	gtctgctgtt	ttatcgagaa	agtactacta	actttgtaag	agcctggttc	6960
cctaaccttc	cctctggctc	tcaagaaatg	tccacaacca	ctggcggtga	cattagtggt	7020
aactatggtt	atattgataa	caaacatagt	gacgatgttc	catttaaaca	atatttctat	7080
atggatgacc	acggtggtat	tgacactgat	gtttcaggga	tattatctat	taatacgaac	7140
attaatcatt	caaaagttaa	agtaatagtg	aaagccgaag	gtatcacaga	gcaaactttt	7200
gtagcgagcg	aaaacagtaa	tgtccccacc	aatccgtccc	gcttcgaaga	aatgaattat	7260
cagtttaaag	agcttgaaat	agatatetee	acactgacat	ttcataataa	tgaagcaagt	7320
attgatatca	cctttatcgc	atttgctgag	aaatttgacg	ataatagtaa	tgatcgtaac	7380
ttaggcgaag	aacatttcag	tattcgtatt	atcaaaaaag	cggaaactga	taatgccctg	7440
accctgcacc	ataatgcaaa	cggggcgcaa	tatatgcagt	ggggaaactc	ttgtattcgc	7500
cttaatacgo	: tatttgcccg	tcaattaatt	agccgagcca	acgcggggat	agatactatt	7560
ttgagca tg g	acactcagaa	tattcaggaa	cctaaattag	gagaagattc	tcctgatgct	7620
atggaaccaa	tggacttcaa	cggcgccaac	agcctctatt	tctgggaact	gttctactac	7680
accccgatgo	tgattgctca	acgtttgctg	cacgaacaaa	acttcgatga	ggctaaccgt	7740
tggctgaaat	atgtctggaa	cccatccggt	tatattgtca	atggtcaaat	gcaacattac	7800
cgctggaatg	g ttcgcccatt	. acaagaagac	actagttgga	acgatgatcc	gttggattca	7860
tttgatcctq	g ataccatago	: tcaacatgat	. ccaatgcact	. acaaagtege	cacctttatg	7920
cgcaccctag	g atctgttgat	. cgaacgggga	gattacgcct	atcgccaatt	. ggagcgggac	7980
acactcgct	g aagccaaaat	gtggtatatg	g caggcactgo	c atctattggg	tgataaacct	8040
catctacca	c tcagttcago	atggaatgat	ccagagctag	g aagaggccgc	agctcttgaa	8100
aaacaacag	g cacatgccaa	a agaaatagca	a gatttacgad	c aaggacttco	tacatccaca	8160
gggtctaaa	g atgaaatcaa	a aacagatctt	ttcctgccg	c aagtcaacga	agtgatgctg	8220
agctactgg	c agaaactaga	a acaacggttg	g tataacctgo	c gccataacct	ctctattgat	8280
ggtcaacct	t tacatttgco	c tattttcgca	a acaccagca	g atccaaaago	gctgctcago	8340
gccgctgtc	g ccagttcaca	a aggtggaagt	aatcttcca	t cagaatttat	atcagtgtgg	8400
cgtttccct	c atatgctgga	a aaacgcccgt	agtatggtc	a gtcagctaad	c ccaattcggc	8460
tccacattg	c aaaatatta	t cgaacgtca	a gatgcggag	g cattaaacad	c gctgttgcaa	a 8520
aatcaggcg	g cagaactga	t attgaccaa	t ctcagcata	c aggacaaaa	catccaagag	g 8580
ctggatgct	g aaaaaactg	t gctagaaaa	a aaccgcgcc	g gaacccagte	c gcgttttgat	8640
agctacago	a aattctacg	a tgaagacat	c aacgcgggt	g aaaaacagg	c aatggcgtt	g 8700

cgtgcttccg	tcgctggcat	ctctacagcc	cttcaagcat	cacatctggc	gggcgcagca	8760
cttgatctgg	cgcccaacat	cttcggcttc	gctgatggtg	gcagccgttg	gggggcgatc	8820
gcccaagcca	caggtaatgt	catggagttc	tccgccagtg	ttatgaacac	cgaagcggat	8880
aaaatcagcc	aatctgaagc	ctaccgtcgg	cgtcgtcagg	aatgggaaat	tcagcgtaat	8940
aacgccgagg	cagagctgaa	acaaatcgat	gctcaacttg	gttcgctggc	agtgcgccgt	9000
gaagccgcag	tattgcagaa	aaccagccta	aaaacccaac	aagagcagac	tcatgcacaa	9060
ctgaccttcc	tgcaacataa	gttcagtaat	caggcgctgt	acaactggct	gcgtggtcga	9120
ttgtccgcca	tttacttcca	gttctatgat	ttaacggtag	ctcgctgttt	gatggcggaa	9180
atggcctatc	gctgggagac	taacgatacc	gcatcacgct	ttatcaaacc	cggcgcctgg	9240
cagggaaccc	atgccggttt	gctcgcgggt	gaaaccttaa	tgctgaatct	ggcacagatg	9300
gaagatgccc	acctgaaaca	ggataaacgc	gtactggagg	tagaacgtac	cgtttcgctg	9360
gccgaagtct	atgccaaatt	accgcaagat	aaatttatcc	tgactcagga	aatagagaag	9420
ttggtgagta	aaggttcagg	cagggccggc	aaggacaata	ataagctggc	gtttagtacc	9480
aataccaata	cctctctaga	agcgtccatt	tcgttatcta	ccttgaacat	tagcagcgat	9540
tatcctgatt	ctattggtaa	aacccgtcgt	att aaaca ga	tcagcgttac	cctgccagca	9600
ctgctaggac	cctatcagga	tgtgcaagca	attctgtctt	acagcggaaa	agcctctgaa	9660
ttggctgaaa	gttgcaaatc	attagcggtt	tctcatggga	tgaatgacag	cggtcagttc	9720
caactggatt	tcaacgatgg	caaattcctg	ccgttcgaag	gaatcaaaat	cgatgaaggt	9780
acgctgacat	tgagcttccc	aaatgcaatt	agtaaagaag	acaaaaaaga	сдаааааддс	9840
aaacaacaag						
_	ccatgctgga	gagtctgaac	gacatcattc	tgcatattcg	ctacaccatt	9900
					ctacaccatt tgcaaggagt	
cgccaataac	gattttaatt	aagtgctaaa	acaggcccct	aagcggggcc		9960
cgccaataac ctttcatgca	gattttaatt aaattcacaa	aagtgctaaa gatttcagta	acaggeeeet ttacagaaet	aagcggggcc atcattgccc	tgcaaggagt	9960 10020
cgccaataac ctttcatgca gcgctatcac	gattttaatt aaattcacaa gggaatgggg	aagtgctaaa gatttcagta gaagctttaa	acaggeceet ttacagaact eccecacegg	aagcggggcc atcattgccc gccggatggg	tgcaaggagt aaaggaggag	9960 10020 10080
cgccaataac ctttcatgca gcgctatcac tgtctctgcc	gattttaatt aaattcacaa gggaatgggg gttgcctatc	aagtgctaaa gatttcagta gaagctttaa tctgccgggc	acaggeceet ttacagaact eccecacegg geggttatge	aagcggggcc atcattgccc gccggatggg tccgtcactc	tgcaaggagt aaaggaggag atggccgcgc	9960 10020 10080 10140
cgccaataac ctttcatgca gcgctatcac tgtctctgcc acaacagcgg	gatttaatt aaattcacaa gggaatgggg gttgcctatc cgccggtaac	aagtgctaaa gatttcagta gaagctttaa tctgccgggc agcccatttg	acaggecect ttacagaact ecceacegg geggttatge gtetgggetg	aageggggee atcattgeee geeggatggg teegteacte ggattgeaac	tgcaaggagt aaaggaggag atggccgcgc gccttaaact	9960 10020 10080 10140 10200
cgccaataac ctttcatgca gcgctatcac tgtctctgcc acaacagcgg tccgccgccg	gatttaatt aaattcacaa gggaatgggg gttgcctatc cgccggtaac cacccatttt	aagtgctaaa gatttcagta gaagctttaa tctgccgggc agcccatttg ggcgttccac	acaggeceet ttacagaact cccccacegg geggttatge gtetgggetg attatgatga	aagcggggcc atcattgccc gccggatggg tccgtcactc ggattgcaac aaccgatacc	tgcaaggagt aaaggaggag atggccgcgc gccttaaact gttatgacca	9960 10020 10080 10140 10200 10260
cgccaataac ctttcatgca gcgctatcac tgtctctgcc acaacagcgg tccgccgccg	gattttaatt aaattcacaa gggaatgggg gttgcctatc cgccggtaac cacccatttt ggtactggtg	aagtgctaaa gatttcagta gaagctttaa tctgccgggc agcccatttg ggcgttccac gtagcggatc	acaggeceet ttacagaact eccecacegg geggttatge gtetgggetg attatgatga aatecegega	aagcggggcc atcattgccc gccggatggg tccgtcactc ggattgcaac aaccgatacc cgaatcgaca	tgcaaggagt aaaggaggag atggccgcgc gccttaaact gttatgacca tttctggggc	9960 10020 10080 10140 10200 10260 10320
cgccaataac ctttcatgca gcgctatcac tgtctctgcc acaacagcgg tccgccgccg cagatggcga tcaacttagg	gatttaatt aaattcacaa gggaatgggg gttgcctatc cgccggtaac cacccatttt ggtactggtg caccgccttt	aagtgctaaa gatttcagta gaagctttaa tctgccgggc agcccatttg ggcgttccac gtagcggatc accgttaccg	acaggecett ttacagaact cccccaccgg gcggttatgc gtctgggctg attatgatga aateccgcga gataccgttc	aagcggggcc atcattgccc gccggatggg tccgtcactc ggattgcaac aaccgatacc cgaatcgaca ccgtctggag	tgcaaggagt aaaggaggag atggccgcgc gccttaaact gttatgacca tttctggggc ttacagggta	9960 10020 10080 10140 10200 10260 10320 10380
cgccaataac ctttcatgca gcgctatcac tgtctctgcc acaacagcgg tccgccgcgcg cagatggcga tcaacttagg gccgattgga	gattttaatt aaattcacaa gggaatgggg gttgcctatc cgccggtaac cacccatttt ggtactggtg caccgccttt atattggcaa	aagtgctaaa gatttcagta gaagctttaa tctgccgggc agcccatttg ggcgttccac gtagcggatc accgttaccg cccaaggcaa	acaggeceet ttacagaact ceeceacegg geggttatge gtetgggetg attatgatga aatecegega gatacegtte cacccaagac	aagcggggcc atcattgccc gccggatggg tccgtcactc ggattgcaac aaccgatacc cgaatcgaca ccgtctggag aactggcaaa	tgcaaggagt aaaggaggag atggccgcgc gccttaaact gttatgacca tttctggggc ttacagggta agtcatttca	9960 10020 10080 10140 10200 10260 10320 10380 10440
cgccaataac ctttcatgca gcgctatcac tgtctctgcc acaacagcgg tccgccgcg cagatggcga tcaacttagg gccgattgga ggctgatata	gattttaatt aaattcacaa gggaatgggg gttgcctatc cgccggtaac cacccatttt ggtactggtg caccgccttt atattggcaa tagcccagat	aagtgctaaa gatttcagta gaagctttaa tctgccgggc agcccatttg ggcgttccac gtagcggatc accgttaccg cccaaggcaa ggacaagtac	acaggecect ttacagaact cceceacegg geggttatge gtetgggetg attatgatga aatecegega gatacegtte cacceaagae atttactggg	aagcggggcc atcattgccc gccggatggg tccgtcactc ggattgcaac aaccgatacc cgaatcgaca ccgtctggag aactggcaaa taaatcacca	tgcaaggagt aaaggaggag atggccgcgc gccttaaact gttatgacca tttctggggc ttacagggta agtcatttca acagattttt	9960 10020 10080 10140 10200 10260 10320 10380 10440 10500

atgaaattac tetecateea	caggccgccg	cgcaacgtta	tctacacaca	gtgtattacg	10680
gcaaccggac agccagcaaa	acgttacccg	gtctggatgg	cagegeeeca	ccacaagcag	10740
actggttatt ctatctggta	tttgattacg	gcgaacgcag	taacaacctg	agaacgccgc	10800
cagcattttc gactacaggt	agctggcttt	gtcgccagga	ccgtttttcc	cgttatgaat	10860
atggttttga gattcgtacc	cgccgcttat	gccgtcaggt	attgatgtat	caccacctgc	10920
aagctctgga tagcgagata	aaagaacaca	acggaccaac	gctggtttca	cgcctgatac	10980
tcaattatga cgaaagcgca	ategecagea	cgctggtatt	cgttcgtcga	gtaggccacg	11040
agcaagacgg tactgccgtc	accetgeege	cattagaatt	ggcgtatcag	gatttttcac	11100
cgcaacataa cgctcgctgg	caatcgatga	atgtgctggc	aaacttcaat	gccattcagc	11160
gctggcaact agttgatcta	aaaggcgaag	gattccccgg	tctgctatat	caagataaag	11220
gegeetggtg gtaeegetee	gcacaacgtt	ttggcaaaat	tggctcagat	gccgtcactt	11280
gggaaaaaat gcaacctttg	tcggttatcc	cttccttgca	aagtaatgcc	tcgctggtgg	11340
atatcaatgg agacggccaa	cttgactggg	ttatcaccgg	accgggatta	cggggatatc	11400
atagtcagca tccagatggc	agttggacac	gttttacccc	gctcaacgct	ctgccagtgg	11460
aatatactca tecaegegeg	caactcgccg	atttaatggg	agctggactt	tctgatttag	11520
tactgatcgg ccctaagagt	gtacgtttat	atgccaatac	ccgcgacggc	tttgccaaag	11580
gaaaagatgt agtgcaatcc	ggtgatatca	cactgccagt	accgggcgcc	gatccgtgta	11640
agttggtggc atttagtgat	gtattgggtt	ccggtcaggc	acatctggtt	gaagtgagcg	11700
cgactaaagt cacctgctgg	cctaatctgg	ggcacggacg	ttttggtcaa	ccaattactc	11760
ttccgggatt tagccaacca	gaagcgacgt	ttaatcctgc	tcaagtttat	ctggccgatc	11820
tagatggcag cggcccgact	gatctgattt	atgttcacac	agatcgtctg	gatatcttcc	11880
tgaataaaag cggcaacggc	ttcgccgcac	cagtaactct	ccccttccca	gccggagtgc	11940
gttttgatca tacctgtcag	ttacaagtgg	ccgatgtaca	agggttaggc	gtcgccagcc	12000
tgatattaag tgtgccgcat	atgactcccc	atcactggcg	ttgcgatctg	ассаасасаа	12060
aaccgtggtt actcagtgaa	atgaacaaca	atatgggggc	tcatcacacc	ctgcgttacc	12120
gtagttccgc ccagttctgg	ctggatgaaa	aagccacggc	actggatgcc	ggacaaatac	12180
cagtttgtta tctacccttc	ccggtacaca	ccctatggca	aacggaaata	gaggatgaaa	12240
tcagcggcaa caaattagtc	acaatactac	gttatgcaca	tggcgcctgg	gatggacgtg	12300
agcgagaatt tcgcggattt	ggttatgttg	aacagaaaga	cagccatcaa	ctggcccaag	12360
gcagtgcgcc agaatgcaca	ccacctgcac	tgacccaagg	caacgcgcct	gaactcacat	12420
cacccgcgct gacccaaggc	aacgctccag	aactcacacc	acctgcgatg	acccaaagca	12480
acgcgcctga actcacatca	cccgcgctga	cccaaggcaa	cgcgccagaa	ttcacatcac	12540
ccgcgctggc ccaaggcaat	gcgccagaac	tcacaccacc	tgcgatgacc	aaaaactggt	12600

atgccaccgg	aatacccatg	atagataaca	cattatcgac	agagtattgg	catggtgatc	12660
accaagcttt	tgccggtttt	tcaccacgct	ttacgacctg	gcaagatggt	caagatattc	12720
tgctcacacc	ggaaaatgat	aacagtcagt	actggctaaa	ccgggcactg	aaaggtcaac	12780
tgctacgcag	tgaactgtac	ggcgaggatg	gcagtacaca	ggaaaaaatt	ccctacacag	12840
tcactgaatt	tcgcccacag	gtacgtcggt	tacagcatac	cgatagccga	taccttgtgc	12900
tttggtcatc	tgtagttgaa	agccgcaact	atcattacga	acgtatcgcc	agcgatcctc	12960
aatgcagcca	aaagattacg	ctatccagcg	atctatttgg	tcaaccgcta	aaacaggttt	13020
cggtacagta	tccacgccgc	cagcaaccgg	caagcagtcc	gtatcctgat	acgttgcctg	13080
ataagttatt	tgctaacagc	tatgatgacc	agcaacacaa	attacggctc	acctatcaac	13140
agttcagttg	gcatcatctg	accgacaata	ccattctgat	gttaggatta	ccggatagta	13200
cccgcagcga	tatctttgct	tatagcgctg	aacatgtccc	tactggtggt	ctaaatctgg	13260
aaatcctaaa	tgataaaaat	agtctgattg	cggagaataa	acctcgtgaa	tacctcggcc	13320
agcaaaaaac	cgtttatacc	gacgggcaaa	atgcaacgcc	atcgcaaacg	ccaacacgac	13380
aagcgctgat	tgccttcacc	gagacaacag	tatttaatca	atccacacta	tcagcgtttg	13440
atgggagtat	ctcatctgct	caattgtcaa	cgacgctgga	acaagccgga	taccagcaaa	13500
cagattatct	attcccgcgc	actggagaag	ataaagtctg	ggcagctcgt	cgtggctata	13560
ctgattacgg	cacageegaa	cagttctggc	ggccgcaaaa	acagagcaac	actcaactca	13620
cgggcaaaat	cacgctcact	tgggatgcaa	actattgcgt	cgtcacacaa	acccgggatg	13680
cggctggact	gacaacctca	gccagatatg	attggcgttt	tctgaccccc	gttcaactca	13740
cggatatcaa	cgacaatcag	caccttacca	cgctggatgc	actgggccga	ccaatcacac	13800
tgcgcttttg	gggaaccgaa	aacggtaaga	tgactggtta	ttcttcaccg	gaaaaaatat	13860
cgttttctcc	accatctgat	gttgacgccg	cgattaagtt	aacaacgcca	atccctgtag	13920
cacagtgtca	ggtctacgca	cccgaaagct	ggatgcccat	attaaagaaa	accctcaata	13980
acctggcaga	gcaagagcgg	aaagagttat	ataacacccg	aatcatcacc	gaagacggac	14040
gcatctgtac	cctagctcac	cgccgctggg	taaaaagcca	aagtgcagtc	acccagccaa	14100
tcaatctgto	aaacggcagt	ccccgtttac	cccctcatag	cctcacattg	actacggatc	14160
gttatgaccg	cgatcttaag	caacagatto	gtcaacaagt	agtattcagt	gatggctttg	14220
gccgtttact	gcaagcatct	gtacgacatg	aagcaggcga	agcctggcaa	cgtaaccaag	14280
acggcgctct	ggtgacaaaa	atggaagata	ccaaaacgcg	ctgggcggtt	acgggacgca	14340
ctgaatatga	caataaggga	caaccgatac	gcacctatca	accctatttc	ctcaacgact	14400
ggcaatacgt	cagtaatgac	agtgcccggc	ggacagaaga	agcctatgca	gatacccatg	14460
tctatgatco	cattggtcga	gaaatcaagg	tcactaccgc	: aaaaggctgg	ttccgtcgaa	14520

WO 99/42589 PCT/EP99/01015 - 30 -

ccttgttcac tccttggttt actgtcaatg aagatgaaaa tgacacagct actgaggtga	14580
aggtaaagaa gaaagaatgt aaagaaggta aagaaggtaa agatgtaatt tgatcaatcc	14640
cgcccggttg aagggcggga aacataacat aatatagagg tgaaacgtgt cattcataat	14700
gccgtcagat actcaactta tgagttggtt gatcattggt tttattgcgg cctggggcgg	14760
attagtaagg tacctcattg atatacaaaa caaacaatgt aaatggaatt ggatcaacgt	14820
actctgtcaa ctcattatct cctgttttac cggtatattg ggaggactgc tgagttttga	14880
aagcggcggc agcccctata tgacttttgc gattgccggg ctatttggca ccacgggaag	14940
ttctggattg aactggatct ggcgtcgcct ttttatgcat tatcgcgatg atggaggaaa	15000
gcaataaggc atteccactg ccgcaaaaac catetgtete cggcagttaa accgggaaat	15060
tacctactac aactattgta agaaaacgaa tatatagaaa aactaacatg cagataaaaa	15120
ctgcgattgc agaacagatg acacacaacg ccccaacaac gaggtaaatc atg aaa Met Lys 1	15176
aac atc gat cct aaa ctt tat caa aag acc cct gtc gtc aac atc tac Asn Ile Asp Pro Lys Leu Tyr Gln Lys Thr Pro Val Val Asn Ile Tyr 5 10 15	15224
gat aac cga ggt cta acg atc cgt aac atc gac ttt cac cgt acc acc Asp Asn Arg Gly Leu Thr Ile Arg Asn Ile Asp Phe His Arg Thr Thr 20 25 30	15272
gca aac ggc gat acc gat atc cgt att act cgc cat caa tat gac tcc Ala Asn Gly Asp Thr Asp Ile Arg Ile Thr Arg His Gln Tyr Asp Ser 35 40 45 50	15320
ctt ggg cac cta agc caa agc acc gat ccg cgt cta tat gaa gcc aaa Leu Gly His Leu Ser Gln Ser Thr Asp Pro Arg Leu Tyr Glu Ala Lys 55 60 65	15368
caa aaa tot aac ttt oto tgg cag tat gat ttg acc ggt aat att ttg Gln Lys Ser Asn Phe Leu Trp Gln Tyr Asp Leu Thr Gly Asn Ile Leu 70 75 80	15416
tgt aca gaa agc gtc gat gct ggt cgc act gtc acc ttg aat gat att Cys Thr Glu Ser Val Asp Ala Gly Arg Thr Val Thr Leu Asn Asp Ile 85 90 95	15464
gaa ggc cgt ccg cta ctg aca gta act gca aca ggt gtc ata caa acc Glu Gly Arg Pro Leu Leu Thr Val Thr Ala Thr Gly Val Ile Gln Thr 100 105 110	15512
cga caa tat gaa acg tct tcc cta ccc ggt cgt ctg ttg tct gtt acc Arg Gln Tyr Glu Thr Ser Ser Leu Pro Gly Arg Leu Leu Ser Val Thr 115 120 125 130	15560
gaa caa ata cca gaa aaa aca tcc cgt atc acc gaa cgc ctg att tgg Glu Gln Ile Pro Glu Lys Thr Ser Arg Ile Thr Glu Arg Leu Ile Trp 135 140 145	15608
get gge aat age gaa gea gag aaa aac eat aat ett gee age eag tge Ala Gly Asn Ser Glu Ala Glu Lys Asn His Asn Leu Ala Ser Gln Cys 150 155 160	15656
gtg cgc cac tat gac acg gcg gga gtc acc cga tta gag agt ttg tca	15704

Val Arg His '	Tyr Asp	Thr Ala	Gly '	Val '	Thr .	Arg		Glu 175	Ser	Leu	Ser	
ctg acc ggt Leu Thr Gly 180			Gln :									15752
caa gaa gct Gln Glu Ala 195					Glu							15800
ctg gct gat Leu Ala Asp												15848
gct tta ctc Ala Leu Leu												15896
tat gat gtg Tyr Asp Val 245		_			_						_	15944
caa ccg gaa Gln Pro Glu 260			Arg									15992
caa aaa tta Gln Lys Leu 275												16040
tat gaa ccg Tyr Glu Pro		Gln Gli									Pro	16088
tca gat gcc Ser Asp Ala			-		-				_	_		16136
ggc aat gtc Gly Asn Val 325									Arg			16184
cac aat cag His Asn Gln 340			o Glu					Tyr				16232
tat cag ctt Tyr Gln Leu 355							Ala					16280
caa agt aac Gln Ser Asn	caa ctt Gln Let 375	ı Pro Se	c ctc r Leu	acc Thr	cta Leu 380	Pro	tct Ser	gat Asp	aac Asn	aac Asr 385	Thr	16328
tac acc aac Tyr Thr Asn					Tyr					' Asr		16376
act aaa atc Thr Lys Ile 405	Gln Hi			Ala					туг			16424
aac atc acg Asn Ile Thr												16472

420	425		430	
acc gaa gat ccg Thr Glu Asp Pro 435	gcg caa gta g Ala Gln Val A 440	at gct tta ttt sp Ala Leu Phe 445	gat gca ggc gga ca Asp Ala Gly Gly Hi	at 16520 is 50
			tgg aat aca cgc gg Trp Asn Thr Arg Gl 465	
			e aag ggc gcc aat ga D Lys Gly Ala Asn As 480	
gat cgg gaa tgg Asp Arg Glu Trp 485	Tyr Arg Tyr S	agt agt gac ggg Ser Ser Asp Gly 190	g aga agg ata tta ag y Arg Arg Ile Leu Ly 495	aa 16664 ys
			a aca cag aga ata ad n Thr Gln Arg Ile Th 510	
			a aac agc acg atc a n Asn Ser Thr Ile T 5	
			a geg ggt egg gea e 1 Ala Gly Arg Ala G 545	
			g gaa gat atc gac a o Glu Asp Ile Asp A 560	
-	Tyr Ser Tyr A		c ggt tcc agt caa c e Gly Ser Ser Gln L 575	
			g gaa gag tac tat c u Glu Glu Tyr Tyr P 590	
			g cgg aca gaa gcc a s Arg Thr Glu Ala S 5	
tat aaa acc atc Tyr Lys Thr Ile	cgt tat tca c Arg Tyr Ser (615	ggt aaa gag cgg Gly Lys Glu Arg 620	g gat gcc acc gga c g Asp Ala Thr Gly L 625	ta 17048 eu
	Tyr Arg Tyr 1		g gta gga cga tgg t p Val Gly Arg Trp L 640	
	Ala Gly Thr V		g aat tta tat cgg a u Asn Leu Tyr Arg M 655	
			t gat gga tta atg c o Asp Gly Leu Met F 670	
			a aat aaa gta gca g s Asn Lys Val Ala A 5	

								aca Thr								17288
_		_						tta Leu 715					_	_	_	17336
								tct Ser								17384
								ctg Leu								17432
								agc Ser								17480
								aag Lys								17528
								cag Gln 795								17576
aaa Lys	gcc Ala	tgg Trp 805	act Thr	cct Pro	cta Leu	gat Asp	act Thr 810	aag Lys	atg Met	gca Ala	agg Arg	cag Gln 815	ttt Phe	gct Ala	agt Ser	17624
								tct Ser								17672
								aaa Lys							tca Ser 850	17720
					Thr			aaa Lys								17768
				Glu										Leu	cat His	17816
gaa Glu	att Ile	aat Asn 885	Met	gat Asp	ctt Leu	tat Tyr	gag Glu 890	Phe	acc Thr	att	gac Asp	gga Gly 895	caa Gln	aag Lys	cta Leu	17864
		Leu					Ser					Pro			tta Leu	17912
	Asp					Glu					: Ile				cat His 930	17960
					Āla					Leu					ctt Leu	18008

aaa aat gta aaa cct tat aag aga taa cgaaaaatta atattcttta 18 Lys Asn Val Lys Pro Tyr Lys Arg 950 955	8055
totactitta atagocotot tgaacttaca otoaaggggg ggaaaccaaa taagaaacca 18	8115
tetttaataa caageeatga aagaatattt attteatgge ttgattaett ttaaeattea 18	8175
atattaaata attaaaacaa tatctaacca attaaaataa caatacctta tttatcatat 1	.8235
taaaatatca aatcagaaat taatgaattt aagggttett tatatttatt tetgagagea 1	.8295
taggcacaat accttaccga tggcgctgga cgtgattcaa aatccagaaa tgctatattt 1	.8355
tcatcaatat gggcagaata gcgcatttca ttgggagtca ttaaacttat cgcgacaccc 1	.8415
gettttacca gatecaatet attagtaaaa teagggaeeg teaataaege taaattttgg 1	.8475
tattcaggga gataattcaa tggcataaaa ttattgcatt gttttaaaaa agcactatta 1	L 8535
tgctgaacaa aaggaaaact agatattatt tcatcagcgt gactttctgg ttctaaaata 1	L8 59 5
tcatgggata cagcaagaga cagcatttga taagcaccat ctatcctgat gatatcatca 1	L8655
ttatctggat aacattcagt cgtcacataa actgttatat cccctttcat taaggaagaa 1	18715
aataccgcat cttgccttat taaatcatca attagaaaat tgttgattat acaaatatcg 1	18775
cgataatgat aacgttgcac cgctcttttt acgaccgtag atattttatt aacatattct 1	18835
ccacttgtgc caataaccag tttgtctctg tttgataatt tataatttct acgacaattc 1	188 9 5
caattattet caactttcag gateetttca taacaeggea geaactettg atatagtgee 1	1895 5
tttccctctt ctgtgagctt ggtttttccc ggtagtcgct caaacaattg acaccccaca 1	19015
cgctgttcca gttgatatac gagcctgcta agtggagaag gggtaataca aagcgtatcc	19075
gccgctaacg tgaatgactc tttcttagct gattccataa aatactttag ttgctttgaa :	19135
caaaatatca tcacataccc tcttgttttc attccagaaa tagaatatta accatagaac :	19195
atgacaacga tgtttctact ttgcattctt ttacattagg acatgcgtta atggacattg	19255
aatttcacta catcaattgt taatatttat ttaatacttg cacaataatt ataaaataaa	19315
tataacttag ttaattattt cttgatattg atcatggtaa gttttcctca atacctacag	19375
aagtagatat tattttatet teeagtaate tategtttgg egaeggaggt egattettee	19435
attgggatat tcaacccatt cgccgccttt cttattaatt acagtgattt ttggcatttt	19495
ggtttcatcc aacttaggtt tataggtgat tttccattta gcacccggtg ttaacttcaa	19555
cctaaaggga tacataccaa cttcaccttg taagaatatt ctgtttggtc taccttcaac	19615
gactttcaaa atggggtaaa taaccgggct aaaatcaatc gtatccaatg catcaatttc	19675
gctgatattt gtccgggctg catcattgat aaatgcgatt aaatcggttg ctgaatacgg	19735
aatagcatct ttcactagat gacggacatc ggtataactc actgacacaa aggctcggtc	19 79 5
aatcttccac ttacatcgac cgccaccatt aaaaggtagt tttgcctgaa agtaaccggt	19855
tttcggatca gcttttacat ccagacgtaa tccgttataa gttggtacct taaaaggcga	19915

catattggaa	tctaaacgat	atttaaggca	atcttttgag	atatacacag	cggatacatg	19975
cggctgtgtg	tatttaggtg	cgactccttc	tacagtaatc	cactgattct	ctttgggagg	20035
agagagcggc	tcatttgggt	cagcacagcc	tgatattaaa	atcacggata	agacagataa	20095
gtatttcttg	atatttatca	tggtaagttt	tecteaacte	ctacagcgtt	atctgcatgt	20155
gtgtccaatt	ccagatette	ctgtttatct	atttagaaat	aaataagcta	cgctgatagc	20215
attacttcat	atttccatac	atgaatcgaa	aatcgacttc	ttgagtgccg	ttatcaattt	20275
tgccgcccgg	atattcaacc	cactcgccgc	ctttcttatt	agtcaccgtg	accttcgcca	20335
ttttggtttc	atccagctta	ggcttaaaaa	taattttcca	tttagctcct	ggagttaacg	20395
tgagttgaaa	aggacgcatt	tttaatactt	caccttgtaa	gaatattctg	ttcgggcgac	20455
cttcaacgac	tttcaaaaca	gggtaaataa	ccgggctaaa	atcaatcgta	ttcaatgtcg	20515
agattttgct	aatattcatc	tggactatgc	cattgataga	tgcgattaaa	ccggttgctg	20575
aatacggaat	agcatctttc	accagatggc	tgacatcagt	ataactcacc	gatacaaagg	20635
cccggttaat	tttccattta	catcgtcccc	ctccattaaa	aggtagtttt	gcttgaaaat	20695
aaccggtttg	tggatcggcc	ttcactttca	gacgaagccc	attataggtc	ggcactttaa	20755
aaggcgacat	attggaatcc	agacgatact	caaggcaatc	ctttgatatg	tattctgcgg	20815
atacatgtgg	ttcggtatat	ttcggcgcta	ccccttctac	cgtgatccat	tgattttctt	20875
taggagggga	aagcggctca	tttgggtcag	cacagcctga	tattaaaatc	actgacaaga	20935
caaataagta	ttttttaaca	tttatcatgg	taagttttcc	tcaattccta	cagcattatc	20995
cgcataaata	tcctgtcaag	aatagcgttc	attgatttcg	tcaccaaaga	aacaagatag	21055
taaaaatcct	attaccacag	ataaaaaaca	ccgcttatgc	cgtgagtaat	agtgagttga	21115
gcgacaggga	tacagcagtg	catccccatc	aattagtccc	tttgaataaa	gggaacagaa	21175
tttgaaattt	ccgtcatacc	gtccatatta	cggaacttag	attatgatta	ttaaatcacc	21235
accaaatggc	aagaaaaatt	ttcatttttt	aatttacgaa	gaatgaattt	gtaagaaagt	21295
gttacaaact	taatagaaat	taatttactg	ttaatctaat	gaaggatgaa	attataaaaa	21355
taacccattt	ctcagggaca	acaatccaca	atatatagaa	ccactggtcc	tcacttaatt	21415
tcctgtcagg	agtagaaata	tcctgatgac	tcagtcgatg	acatacagca	atgtcattgg	21475
tattgagact	accgactgtt	taataaattt	cttttgtctt	taatggcgag	atacaagtga	21535
ttcactattt	aagcactatc	gataaataag	attccaaaat	agcgccatat	cttacaccac	21595
tcataattct	atgtataaca	attggttaaa	taggatcatg	tgtaacagga	ttatgaaacg	21655
ttatttatat	caaatctatc	aattatttta	tatatagttt	cacagtcaca	ctcgctatct	21715
ggtaccttca	taaccaactg	ccctccctgc	gctaccttct	gataacaaca	gctacactaa	21775
ctatacccgc	gcctataatt	atgaccgtgt	gaaaattcag	cgtagttcac	cggccacgca	21835

aaataactac	acqaaatatt	gctccccaga	gaaacaccgt	tcgaggttgt	ttcaatgaaa	21895
			caagatatta			
			ttcacgtcat			
			aaaatcccct			
			tttatttcat			
			taatteetaa			
			gagagtataa			
			aatgccttta			
			gatacctgat			
			attctggtat			
			actattatgc			
			taaaatatca			
			atcatcgttg			
			agaagaaaat			
			aatatcacga			
			atatteteca			
			ataatgccaa			
			taatgeette			
			geceacaegt			
			g cgtatccgcg			
			ctttgaacaa			
			a gaatggaatg			
			aaagtatgca			
tcatctttat	tcttaacagt	tatcaaacaa	a ttttcattat	tattgcaaaa	taaatacaac	23275
cccttcttat	gttacaataa	a tgattataa	a gaaatttcad	c atattatcat	taagtaataa	23335
tgggcacaat	taaccattt	a attaaacat	t tcaattggt!	gacaaagact	cattatgtto	23395
aacatgtaat	gagcgcaati	ttaacatta	a ataaattac	a tagttcatat	tcattatcac	23455
tgagatcago	tttt ttc gta	a tagtacatca	a tgtgaacaa	t accgtgccat	t ttcctgccaa	a 23515
atcttatta	a aaagtcagt	t gcaaatttt	g catctgctt	t ttttgcaaca	a gctatttaaa	a 23575
gaaaacagt	g agatagtga	t tatccgaga	g atcaagata	t gtetgetet!	t tacgcacaaa	a 23635
ctgcaaacca	a tttctatgc	a tatctcagc	t atttctcaa	a acctgtatt!	t aatcatctct	23695
tattccgate	g gaacggaat	c attctctga	t tgattcatg	a tgtaaagaca	a atatggatgi	23755
ttcatttac	t tt atg at	t tta aaa g	ga ata aat .	atg aat tcg	cct gta aaa	a 23806

Met	Ile	Leu	Lys	Gly	Ile	Asn	Met	Asn	Ser	Pro	Val	Lys
				960					965			-

gag ata cct gat gta tta aaa atc cag tgt ggt ttt cag tgt ctg aca Glu Ile Pro Asp Val Leu Lys Ile Gln Cys Gly Phe Gln Cys Leu Thr 970 975 980	23854
gat att agc cac agc tct ttt aac gaa ttt cac cag caa gta tcc gaa Asp Ile Ser His Ser Ser Phe Asn Glu Phe His Gln Gln Val Ser Glu 985 990 995 1000	23902
cac ctc tcc tgg tcc gaa gca cac gac tta tat cat gat gca caa cag His Leu Ser Trp Ser Glu Ala His Asp Leu Tyr His Asp Ala Gln Gln 1005 1010 1015	23950
gcc caa aag gat aat cgg ctg tat gaa gcg cgt att ctt aaa cgc acg Ala Gln Lys Asp Asn Arg Leu Tyr Glu Ala Arg Ile Leu Lys Arg Thr 1020 1025 1030	23998
aat oot caa tta caa aat got gta cat ott god atc gta gog oot aat Asn Pro Gln Leu Gln Asn Ala Val His Leu Ala Ile Val Ala Pro Asn 1035 1040 1045	24046
gct gaa ctg ata ggc tat aac aac caa ttt agc ggc agg gcc agt caa Ala Glu Leu Ile Gly Tyr Asn Asn Gln Phe Ser Gly Arg Ala Ser Gln 1050 1060	24094
tat gtc gcg ccg ggt acc gtt tcc tcc atg ttc tcc ccc gcc gct tat Tyr Val Ala Pro Gly Thr Val Ser Ser Met Phe Ser Pro Ala Ala Tyr 1065 1070 1075 1080	24142
ttg act gag ctt tat cgt gaa gca cgc aat tta cac gcc agc gat tcc Leu Thr Glu Leu Tyr Arg Glu Ala Arg Asn Leu His Ala Ser Asp Ser 1085 1090 1095	24190
gtt tat cgc ctg gat act cgc cgc cca gat ctc aaa tca atg gcg ctc Val Tyr Arg Leu Asp Thr Arg Arg Pro Asp Leu Lys Ser Met Ala Leu 1100 1105 1110	24238
agt caa caa aat atg gat acg gaa ctt tcc act ctc tct tta tcc aat Ser Gln Gln Asn Met Asp Thr Glu Leu Ser Thr Leu Ser Leu Ser Asn 1115 1120 1125	24286
gag cta tta ttg gaa agc att aaa act gag tct aag ctg gat aat tat Glu Leu Leu Glu Ser Ile Lys Thr Glu Ser Lys Leu Asp Asn Tyr 1130 1140	24334
act caa gtg atg gaa atg ctc tcc gct ttc cgt cct tcc ggc gcg acg Thr Gln Val Met Glu Met Leu Ser Ala Phe Arg Pro Ser Gly Ala Thr 1145 1150 1155 1160	24382
cct tat cac gat gct tac gaa aat gtg cgt aaa gtt atc cag cta caa Pro Tyr His Asp Ala Tyr Glu Asn Val Arg Lys Val Ile Gln Leu Gln 1165 1170 1175	24430
gat cet ggg ett gag eaa tta aat get tea eea gee att gee ggg etg Asp Pro Gly Leu Glu Gln Leu Asn Ala Ser Pro Ala Ile Ala Gly Leu 1180 1185 1190	24478
atg cat caa gct tcc cta tta ggt att aac gct tca atc tca cct gag Met His Gln Ala Ser Leu Leu Gly Ile Asn Ala Ser Ile Ser Pro Glu 1195 1200 1205	24526
ttg ttt aat att ctg acg gag gag att act gaa ggt aat gct gag gaa Leu Phe Asn Ile Leu Thr Glu Glu Ile Thr Glu Gly Asn Ala Glu Glu	24574

1210	1215	1220	
ctt tat aag aaa Leu Tyr Lys Lys 1225	aat ttt ggt aa Asn Phe Gly As 1230	at atc gaa ccg gct tca ctg gct sn Ile Glu Pro Ala Ser Leu Ala 1235 l	atg 24622 Met 240
		ac aat tta agt gat gaa gaa ctc yr Asn Leu Ser Asp Glu Glu Leu 1250 1255	
	Lys Ala Ser As	at ttc ggc caa caa gaa tat agt sn Phe Gly Gln Gln Glu Tyr Ser 1265 1270	
		cc aac agc aat gat ggc aca gtc al Asn Ser Asn Asp Gly Thr Val 30 1285	
gta tat cga att Val Tyr Arg Ile 1290	acc cgc gaa ta Thr Arg Glu Ty 1295	at aca aca aat gcc aat caa gta yr Thr Thr Asn Ala Asn Gln Val 1300	gac 24814 Asp
gtg gag ctg tt Val Glu Leu Phe 1305	ccc tac ggt g Pro Tyr Gly G 1310	ga gaa aat tat cag tta aat tac ly Glu Asn Tyr Gln Leu Asn Tyr 1315	aaa 24862 Lys 1320
		tc tcc tat tta tcc atc aaa tta al Ser Tyr Leu Ser Ile Lys Leu 1330 1335	
	ı Leu Ile Arg I	tt gaa gga gcg cct cag gtc aac le Glu Gly Ala Pro Gln Val Asn 1345 1350	
gaa tat tca ga Glu Tyr Ser Gl 1355	u His Ile Thr L	ta agt aca act gat atc agt caa eu Ser Thr Thr Asp Ile Ser Gln 60 1365	cct 25006 Pro
		ita tat oct tot agt tot tgg goa 'al Tyr Pro Ser Ser Ser Trp Ala 1380	
		ag gaa tat aac caa tac tct ttc Glu Glu Tyr Asn Gln Tyr Ser Phe 1395	
	_	egt cta tet egt geg aca gaa tta Arg Leu Ser Arg Ala Thr Glu Leu 1410 1415	Ser
	u Glu Ser Ile V	gtg cgt agt gtt aat cag caa ctg Val Arg Ser Val Asn Gln Gln Leu 1425 1430	
	u Val Leu Gly I	aaa gtt ttt ctg act aaa tat tat Lys Val Phe Leu Thr Lys Tyr Tyr 140 1445	
		gaa act gcc cta ata cta tgc aat Glu Thr Ala Leu Ile Leu Cys Asn 1460	
		gat aat caa cct agc caa ttt gat Asp Asn Gln Pro Ser Gln Phe Asp 1475	

ctg ttt aat acg cca tta ctg aac ggc caa tat ttt tct acc gga gat 2 Leu Phe Asn Thr Pro Leu Leu Asn Gly Gln Tyr Phe Ser Thr Gly Asp 1485 1490 1495	:5390
gaa gag att gat tta aat cca ggt agt act ggc gat tgg cgt aaa tcc 2 Glu Glu Ile Asp Leu Asn Pro Gly Ser Thr Gly Asp Trp Arg Lys Ser 1500 1505 1510	:5438
gtg ctt aaa cgt gca ttt aat atc gat gat att tcc ctc tac cgc ctg 2 Val Leu Lys Arg Ala Phe Asn Ile Asp Asp Ile Ser Leu Tyr Arg Leu 1515 1520 1525	5486
ctt aaa att acc aac cat aat aat caa gat gga aag att aaa aat aac 2 Leu Lys Ile Thr Asn His Asn Asn Gln Asp Gly Lys Ile Lys Asn Asn 1530 1535 1540	5534
tta aat aat ctt tct gat tta tat att ggg aaa tta ctg gca gaa att 2 Leu Asn Asn Leu Ser Asp Leu Tyr Ile Gly Lys Leu Leu Ala Glu Ile 1545 1550 1560	:5582
cat caa tta acc att gat gaa ttg gat tta ttg ctg gtt gcc gtg ggt 2 His Gln Leu Thr Ile Asp Glu Leu Asp Leu Leu Leu Val Ala Val Gly 1565 1570 1575	5630
gaa gga gaa act aat tta tcc gct atc agt gat aaa caa ctg gcg gca 2 Glu Gly Glu Thr Asn Leu Ser Ala Ile Ser Asp Lys Gln Leu Ala Ala 1580 1585 1590	25678
ctg atc aga aaa ctc aat acc att acc gtc tgg cta cag aca cag aag 2 Leu Ile Arg Lys Leu Asn Thr Ile Thr Val Trp Leu Gln Thr Gln Lys 1595 1600 1605	25726
tgg agt gcg ttc caa tta ttt gtt atg act tcc acc agc tat aac aaa 2 Trp Ser Ala Phe Gln Leu Phe Val Met Thr Ser Thr Ser Tyr Asn Lys 1610 1620	5774
acg ctg acg cct gaa att aag aat ctg ctg gat acc gtc tac cac ggt 2 Thr Leu Thr Pro Glu Ile Lys Asn Leu Leu Asp Thr Val Tyr His Gly 1625 1630 1635 1640	:5822
tta caa ggc ttt gat aaa gac aag gca aat tta ctg cat gtt atg gcg 2 Leu Gln Gly Phe Asp Lys Asp Lys Ala Asn Leu Leu His Val Met Ala 1645 1650 1655	5870
ccc tat att gcg gcc acc tta caa tta tca tcg gaa aat gtc gcc cat 2 Pro Tyr Ile Ala Ala Thr Leu Gln Leu Ser Ser Glu Asn Val Ala His 1660 1665 1670	5918
tct gtg ctg ctt tgg gca gac aag tta aag ccc ggc gac ggc gca atg 2 Ser Val Leu Leu Trp Ala Asp Lys Leu Lys Pro Gly Asp Gly Ala Met 1675 1680 1685	5966
aca gcc gaa aaa ttc tgg gac tgg ttg aat act caa tat acg cca gat 2 Thr Ala Glu Lys Phe Trp Asp Trp Leu Asn Thr Gln Tyr Thr Pro Asp 1690 1695 1700	86014
tca tcg gaa gta tta gca aca cag gaa cat att gtt cag tat tgt cag 2 Ser Ser Glu Val Leu Ala Thr Gln Glu His Ile Val Gln Tyr Cys Gln 1705 1710 1715 1720	:6062
gcg ttg gcg caa tta gaa atg gtt tac cat tcc acc ggt atc aat gaa 2 Ala Leu Ala Gln Leu Glu Met Val Tyr His Ser Thr Gly Ile Asn Glu	6110

aac gcc ttc cgc ctg ttt gtg aca aaa cca gag atg ttt ggc tcg tca Asn Ala Phe Arg Leu Phe Val Thr Lys Pro Glu Met Phe Gly Ser Ser 1740 1745 1750	26158
act gag gca gta cct gcg cat gat gca ctt tca ctg atc atg ctg acg Thr Glu Ala Val Pro Ala His Asp Ala Leu Ser Leu Ile Met Leu Thr 1755 1760 1765	26206
cgt ttt gca gat tgg gtt aat gcg tta ggc gaa aaa gcc tct tcc gta Arg Phe Ala Asp Trp Val Asn Ala Leu Gly Glu Lys Ala Ser Ser Val 1770 1775 1780	26254
cta gcg gca ttt gaa gct aac agt tta acg gca gaa caa ttg gct gat Leu Ala Ala Phe Glu Ala Asn Ser Leu Thr Ala Glu Gln Leu Ala Asp 1785 1790 1795 1800	26302
gcc atg aat ctt gat gct aat ttg cta ttg caa gcc agt act caa gca Ala Met Asn Leu Asp Ala Asn Leu Leu Gln Ala Ser Thr Gln Ala 1805 1810 1815	26350
caa aac cat caa cat ctt ccc cca gtg acg caa aaa aat gct ttc tcc Gln Asn His Gln His Leu Pro Pro Val Thr Gln Lys Asn Ala Phe Ser 1820 1825 1830	263 9 8
tgt tgg aca tct atc gac act atc ctg caa tgg gtt aat gtt gca caa Cys Trp Thr Ser Ile Asp Thr Ile Leu Gln Trp Val Asn Val Ala Gln 1835 1840 1845	26446
caa ttg aat gtc gcc cca cag gga gtt tcc gct ttg gtc ggg ctg gat Gln Leu Asn Val Ala Pro Gln Gly Val Ser Ala Leu Val Gly Leu Asp 1850 1855 1860	26494
tat att caa tta aat caa aaa atc ccc acc tat gcc cag tgg gaa agt Tyr Ile Gln Leu Asn Gln Lys Ile Pro Thr Tyr Ala Gln Trp Glu Ser 1865 1870 1875 1880	26542
gct ggg gaa ata ttg act gcc gga ttg aat tca caa cag gct gat ata Ala Gly Glu Ile Leu Thr Ala Gly Leu Asn Ser Gln Gln Ala Asp Ile 1885 1890 1895	26590
tta cac gct ttt ttg gac gaa tct cgc agt gcc gca tta agc acc tac Leu His Ala Phe Leu Asp Glu Ser Arg Ser Ala Ala Leu Ser Thr Tyr 1900 1905 1910	26638
tat atc cgt caa gtc gcc aag cca gcg gca gcc ata aaa agc cgt gat Tyr Ile Arg Gln Val Ala Lys Pro Ala Ala Ala Ile Lys Ser Arg Asp 1915 1920 1925	26686
gac ttg tac caa tac tta cta att gat aat cag gtt tcc gct gca atc Asp Leu Tyr Gln Tyr Leu Leu Ile Asp Asn Gln Val Ser Ala Ala Ile 1930 1935 1940	26734
aaa act acc cgg att gcc gaa gcc att gcc agc att caa ctg tac gtc Lys Thr Thr Arg Ile Ala Glu Ala Ile Ala Ser Ile Gln Leu Tyr Val 1945 1950 1955 1960	26782
aac cgc acg ctg gaa aat gta gaa gaa aat gcc cat tca ggg gtt atc Asn Arg Thr Leu Glu Asn Val Glu Glu Asn Ala His Ser Gly Val Ile 1965 1970 1975	26830
agc cgt cag ttc ttt atc gac tgg gac aaa tat aac aaa cgc tac agc Ser Arg Gln Phe Phe Ile Asp Trp Asp Lys Tyr Asn Lys Arg Tyr Ser 1980 1985 1990	26878
acc tgg gcg ggt gtt tct caa tta gtt tac tac ccg gaa aac tat att	26926

Thr Trp Ala Gly Val Ser Gln Leu Val Tyr Tyr Pro Glu Asn Tyr Ile 1995 2000 2005	
gat ccc acc atg cgt atc gga caa acc aaa atg atg gac gca tta ttg Asp Pro Thr Met Arg Ile Gly Gln Thr Lys Met Met Asp Ala Leu Leu 2010 2015 2020	26974
caa tcc gtc agc caa agc caa tta aat gcc gat act gtc gaa gac gcc Gln Ser Val Ser Gln Ser Gln Leu Asn Ala Asp Thr Val Glu Asp Ala 2025 2030 2035 2040	27022
ttt atg tct tat ctg aca tcg ttt gag caa gtg gct aat ctt aaa gtt Phe Met Ser Tyr Leu Thr Ser Phe Glu Gln Val Ala Asn Leu Lys Val 2045 2050 2055	270 7 0
att agc gcg tat cac gat aat att aac aac gat caa ggg ctg acc tat Ile Ser Ala Tyr His Asp Asn Ile Asn Asp Gln Gly Leu Thr Tyr 2060 2065 2070	27118
ttt atc ggc ctc agt gaa act gat acc ggt gaa tac tat tgg cgc agt Phe Ile Gly Leu Ser Glu Thr Asp Thr Gly Glu Tyr Tyr Trp Arg Ser 2075 2080 2085	27166
gtc gat cac agt aaa ttc agc gac ggt aaa ttc gcc gct aat gcc tgg Val Asp His Ser Lys Phe Ser Asp Gly Lys Phe Ala Ala Asn Ala Trp 2090 2095 2100	27214
agt gaa tgg cac aaa att gat tgt cca att aat cct tac cga agc act Ser Glu Trp His Lys Ile Asp Cys Pro Ile Asn Pro Tyr Arg Ser Thr 2105 2110 2115 2120	2 7262
atc cgt cct gtg atg tac aaa tcc cgc ttg tat ctg ctc tgg ttg gaa Ile Arg Pro Val Met Tyr Lys Ser Arg Leu Tyr Leu Leu Trp Leu Glu 2125 2130 2135	27310
caa aag gag atc act aaa caa aca gga aat agc aaa gat ggc tat caa Gln Lys Glu Ile Thr Lys Gln Thr Gly Asn Ser Lys Asp Gly Tyr Gln 2140 2145 2150	27358
acc gag aca gat tat cgt tat gag cta aaa ttg gcg cat atc cgt tat Thr Glu Thr Asp Tyr Arg Tyr Glu Leu Lys Leu Ala His Ile Arg Tyr 2155 2160 2165	27406
gac ggt acc tgg aat acg cca atc act ttt gat gtc aat gaa aaa ata Asp Gly Thr Trp Asn Thr Pro Ile Thr Phe Asp Val Asn Glu Lys Ile 2170 2175 2180	27454
tcc aag cta gaa ctg gca aaa aat aaa gcg cct ggg ctc tat tgt gct Ser Lys Leu Glu Leu Ala Lys Asn Lys Ala Pro Gly Leu Tyr Cys Ala 2185 2190 2195 2200	27502
ggt tat caa ggt gaa gat acg ttg ctg gtt atg ttt tat aac caa caa Gly Tyr Gln Gly Glu Asp Thr Leu Leu Val Met Phe Tyr Asn Gln Gln 2205 2210 2215	27550
gat aca ctc gat agt tat aaa acc gct tca atg caa ggg cta tat atc Asp Thr Leu Asp Ser Tyr Lys Thr Ala Ser Met Gln Gly Leu Tyr Ile 2220 2225 2230	27598
ttt gcc gat atg gaa tat aaa gat atg acc gat gga caa tac aaa tct Phe Ala Asp Met Glu Tyr Lys Asp Met Thr Asp Gly Gln Tyr Lys Ser 2235 2240 2245	27646
tat cgg gac aac agc tat aaa caa ttc gat act aat agt gtc aga aga Tyr Arg Asp Asn Ser Tyr Lys Gln Phe Asp Thr Asn Ser Val Arg Arg	27694

2250	2255	2260		
gtg aat aac cgc Val Asn Asn Arg 2265	tat gca gag gat Tyr Ala Glu Asp 2270	tat gaa att ccc Tyr Glu Ile Pro 2275	tca tcg gta aat Ser Ser Val Asn 2280	27742
Ser Arg Lys Gly	tat gat tgg gga Tyr Asp Trp Gly 2285	gat tat tat ctc Asp Tyr Tyr Leu 2290	agt atg gta tat Ser Met Val Tyr 2295	27790
aac gga gat att Asn Gly Asp Ile 2300	Pro Thr Ile Ser	tac aaa gcc aca Tyr Lys Ala Thr 2305	tca agt gat tta Ser Ser Asp Leu 2310	27838
aaa atc tat atc Lys Ile Tyr Ile 2315	tcg cca aaa tta Ser Pro Lys Leu 2320	aga att att cat Arg Ile Ile His	aat gga tat gaa Asn Gly Tyr Glu 2325	27886
ggg cag caa cgc Gly Gln Gln Arg 2330	aat caa tgc aat Asn Gln Cys Asn 2335	cta atg aat aaa Leu Met Asn Lys 2340	tat ggc aaa cta Tyr Gly Lys Leu	27934
ggt gat aaa ttt Gly Asp Lys Phe 2345	att gtt tat act Ile Val Tyr Thr 2350	agc ttg gga gtt Ser Leu Gly Val 2355	aat cca aat aat Asn Pro Asn Asn 2360	27982
Ser Ser Asn Lys	ctg atg ttt tac Leu Met Phe Tyr 2365	ccc gtt tat caa Pro Val Tyr Gln 2370	tat aac gga aat Tyr Asn Gly Asn 2375	28030
gtc agt ggg ctt Val Ser Gly Leu 2380	Ser Gln Gly Arg	tta cta ttc cac Leu Leu Phe His 2385	cgt gac acc aat Arg Asp Thr Asn 2390	28078
tat tca tct aaa Tyr Ser Ser Lys 2395	gta gaa gct tgg Val Glu Ala Trp 2400	att cct gga gca Ile Pro Gly Ala	gga cgt tct cta Gly Arg Ser Leu 2405	28126
acc aat ccg aat Thr Asn Pro Asn 2410	gct gcc att ggt Ala Ala Ile Gly 2415	gat gat tat gct Asp Asp Tyr Ala 2420	aca gac tcg tta Thr Asp Ser Leu	28174
aac aaa ccg aat Asn Lys Pro Asn 2425	gat ctt aag caa Asp Leu Lys Gln 2430	tac gtc tat atg Tyr Val Tyr Met 2435	act gac agt aaa Thr Asp Ser Lys 2440	28222
Gly Thr Ala Thr	gat gtc tca gga Asp Val Ser Gly 2445	cca gta gat atc Pro Val Asp Ile 2450	aat act gca att Asn Thr Ala Ile 2455	28270
tcc ccg gca aaa Ser Pro Ala Lys 2460	Val Gln Val Thr	gta aaa gcc ggt Val Lys Ala Gly 2465	agc aaa gaa caa Ser Lys Glu Gln 2470	28318
acg ttt acc gcg Thr Phe Thr Ala 2475	gat aaa aat gtc Asp Lys Asn Val 2480	tcc att cag cca Ser Ile Gln Pro	tcc cct agc ttt Ser Pro Ser Phe 2485	28366
gat gaa atg aat Asp Glu Met Asn 2490	tat caa ttt aat Tyr Gln Phe Asn 2495	gct ctc gaa ata Ala Leu Glu Ile 2500	gat ggc tca agt Asp Gly Ser Ser	28414
ctg aat ttt act Leu Asn Phe Thr 2505	aac aat tca gcc Asn Asn Ser Ala 2510	agt att gat att Ser Ile Asp Ile 2515	acc ttt acc gca Thr Phe Thr Ala 2520	28462

ttt gca gag gat gga cgt aaa ctg ggt tat gaa agt ttc agt att cct Phe Ala Glu Asp Gly Arg Lys Leu Gly Tyr Glu Ser Phe Ser Ile Pro 2525 2530 2535	28510
att acc cgc aag gtg agt act gat aat tcc ctg acc ctg cgc cat aat Ile Thr Arg Lys Val Ser Thr Asp Asn Ser Leu Thr Leu Arg His Asn 2540 2545 2550	28558
gaa aat ggt gcg caa tat atg caa tgg gga gtc tat cgc att cgt ctt Glu Asn Gly Ala Gln Tyr Met Gln Trp Gly Val Tyr Arg Ile Arg Leu 2555 2560 2565	28606
aat act tta ttt gct cgc caa tta gtt gcg cga gcc act acc ggt att Asn Thr Leu Phe Ala Arg Gln Leu Val Ala Arg Ala Thr Thr Gly Ile 2570 2580	28654
gat acg att ctg agt atg gaa act cag aat att cag gaa cca cag tta Asp Thr Ile Leu Ser Met Glu Thr Gln Asn Ile Gln Glu Pro Gln Leu 2585 2590 2595 2600	28702
ggc aaa ggt ttc tac gct acg ttc gtg ata cct ccg tat aac cca tca Gly Lys Gly Phe Tyr Ala Thr Phe Val Ile Pro Pro Tyr Asn Pro Ser 2605 2610 2615	28750
act cat ggt gat gaa cgt tgg ttt aag ctt tat atc aaa cat gtt gtt Thr His Gly Asp Glu Arg Trp Phe Lys Leu Tyr Ile Lys His Val Val 2620 2630	28798
gat aat aat toa oat att ato tat toa ggt oag ota aaa gat aca aat Asp Asn Asn Ser His Ile Ile Tyr Ser Gly Gln Leu Lys Asp Thr Asn 2635 2640 2645	28846
ata agc acc acg tta ttt atc cct ctt gat gat gtt cca ttg aac caa Ile Ser Thr Thr Leu Phe Ile Pro Leu Asp Asp Val Pro Leu Asn Gln 2650 2655 2660	28894
gat tac agc gcc aag gtt tac atg acc ttc aag aaa tca cca tca gat Asp Tyr Ser Ala Lys Val Tyr Met Thr Phe Lys Lys Ser Pro Ser Asp 2665 2670 2675 2680	28942
ggt acc tgg tgg ggc cct cac ttt gtt aga gat gat aaa gga ata gta Gly Thr Trp Trp Gly Pro His Phe Val Arg Asp Asp Lys Gly Ile Val 2685 2690 2695	28990
aca ata aac cct aaa tcc att ttg acc cac ttt gag agc gtc aat gtc Thr Ile Asn Pro Lys Ser Ile Leu Thr His Phe Glu Ser Val Asn Val 2700 2705 2710	29038
ctg aat aat att agt agc gaa cca atg gat ttc agc ggc gct aac agc Leu Asn Asn Ile Ser Ser Glu Pro Met Asp Phe Ser Gly Ala Asn Ser 2715 2720 2725	29086
ctc tat ttt tgg gaa ctg ttc tac tat acc ccg atg ctg gtt gcc caa Leu Tyr Phe Trp Glu Leu Phe Tyr Tyr Thr Pro Met Leu Val Ala Gln 2730 2735 2740	29134
cgt ttg ttg cat gag caa aac ttt gat gaa gcg aac cgc tgg ctg aaa Arg Leu Leu His Glu Gln Asn Phe Asp Glu Ala Asn Arg Trp Leu Lys 2745 2750 2750 2760	29182
tat gtc tgg agc cca tcc ggg tat att gtt cac ggc cag att cag aat Tyr Val Trp Ser Pro Ser Gly Tyr Ile Val His Gly Gln Ile Gln Asn 2765 2770 2775	29230

tat caa tgg aac gtc cgc ccg tta ttg gaa gat acc agt tgg aac agt Tyr Gln Trp Asn Val Arg Pro Leu Leu Glu Asp Thr Ser Trp Asn Ser 2780 2785 2790	292 7 8
gat cct ttg gat tcc gtc gat cct gac gcg gta gcg cag cac gat ccg Asp Pro Leu Asp Ser Val Asp Pro Asp Ala Val Ala Gln His Asp Pro 2795 2800 2805	29326
atg cac tat aaa gtt tca acc ttt atg cgc acc ctt gat ctg ttg atc Met His Tyr Lys Val Ser Thr Phe Met Arg Thr Leu Asp Leu Leu Ile 2810 2815 2820	29374
gcg cgc ggc gac cat gct tac cgc caa ttg gag cgc gat acg ctt aac Ala Arg Gly Asp His Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Asn 2825 2830 2835 2840	29422
gaa gcg aag atg tgg tat atg caa gcg ctg cat ctg tta ggc gat aaa Glu Ala Lys Met Trp Tyr Met Gln Ala Leu His Leu Leu Gly Asp Lys 2845 2850 2855	29470
cct tat ctg ccg ctg agt acc aca tgg aat gat cca cga ctg gac aaa Pro Tyr Leu Pro Leu Ser Thr Thr Trp Asn Asp Pro Arg Leu Asp Lys 2860 2865 2870	29518
gcc gcg gat att act acc caa agt gct cat tcc agc tca ata gtc gct Ala Ala Asp Ile Thr Thr Gln Ser Ala His Ser Ser Ser Ile Val Ala 2875 2880 2885	29566
ttg cgg cag agt aca ccg gcg ctt tta tca ttg cgc agc gcc aat acc Leu Arg Gln Ser Thr Pro Ala Leu Leu Ser Leu Arg Ser Ala Asn Thr 2890 2895 2900	29614
ctg acc gat ctc ttc ctg ccg caa atc aat gaa gtg atg atg aat tac Leu Thr Asp Leu Phe Leu Pro Gln Ile Asn Glu Val Met Met Asn Tyr 2905 2910 2915 2920	29662
tgg caa aca tta gct cag aga gta tac aac ctg cgc cac aac ctc tct Trp Gln Thr Leu Ala Gln Arg Val Tyr Asn Leu Arg His Asn Leu Ser 2925 2930 2935	29710
atc gac ggt cag ccg tta tat ctg cca atc tat gcc aca ccg gcg gac Ile Asp Gly Gln Pro Leu Tyr Leu Pro Ile Tyr Ala Thr Pro Ala Asp 2940 2945 2950	29 75 8
ccg aaa gcg tta ctc agc gcc gct gtt gcc act tct caa ggt gga ggc Pro Lys Ala Leu Leu Ser Ala Ala Val Ala Thr Ser Gln Gly Gly Gly 2955 2960 2965	29806
aag ctg ccg gag tca ttt atg tcc ctg tgg cgt ttc ccg cac atg ctg Lys Leu Pro Glu Ser Phe Met Ser Leu Trp Arg Phe Pro His Met Leu 2970 2975 2980	29854
gaa aat gct cgc agc atg gtt agc cag ctc acc caa ttc ggc tcc acg Glu Asn Ala Arg Ser Met Val Ser Gln Leu Thr Gln Phe Gly Ser Thr 2985 2990 2995 3000	29902
tta caa aat att atc gaa cgt cag gac gca gaa gcg ctc aat gcg tta Leu Gln Asn Ile Ile Glu Arg Gln Asp Ala Glu Ala Leu Asn Ala Leu 3005 3010 3015	29950
tta caa aat cag gcc gca gag ctg ata ttg act aac ctg agt att caa Leu Gln Asn Gln Ala Ala Glu Leu Ile Leu Thr Asn Leu Ser Ile Gln 3020 3025 3030	29998
gac aaa acc att gaa gaa ctg gat gcc gag aaa acc gtg ctg gaa aaa	30046

Asp Lys Thr Ile 3035		Asp Ala Glu I 040	Lys Thr Val Leu 3045	Glu Lys	
tcc aaa gcg gga Ser Lys Ala Gly 3050	gca caa teg o Ala Gln Ser A 3055	ege ttt gat a Arg Phe Asp S	agc tat agc aaa Ser Tyr Ser Lys 3060	ctg cat 30094 Leu His	1
gat gaa aac atc Asp Glu Asn Ile 3065		Glu Asn Gln A			2
tcc gca gcc ggg Ser Ala Ala Gly			Ala Ser Arg Leu		D
gca gcg gct gat Ala Ala Ala Asp 3100	ctg gtg cct a Leu Val Pro A	aac atc ttc o Asn Ile Phe (3105	ggc ttc gcc ggt Gly Phe Ala Gly 3110	ggt ggt 3023 Gly Gly	8
agc cgt tgg ggg Ser Arg Trp Gly 3115	Ala Ile Ala				6
tcc gct aat gtt Ser Ala Asn Val 3130	-			-	4
ace tac egt egt Thr Tyr Arg Arg 3145		Glu Trp Glu			2
gaa gcg gag ctg Glu Ala Glu Leu			Leu Lys Ser Leu		.О
cgc cgt gaa gcc Arg Arg Glu Ala 3180	Ala Val Leu				'8
gag cag acc caa Glu Gln Thr Gln 3195	Ala Gln Leu				:6
caa gcg ttg tac Gln Ala Leu Tyr 3210	aac tgg cta Asn Trp Leu 3215	cgt ggc cga Arg Gly Arg	ctg gca gca att Leu Ala Ala Ile 3220	tac ttc 3057 Tyr Phe	14
caa ttc tac gac Gln Phe Tyr Asp 3225		Ala Arg Cys			22
tac cgt tgg gaa Tyr Arg Trp Glu	a att agc gat 1 Ile Ser Asp 3245	gac tct gct Asp Ser Ala 3250	cgc ttt att aaa Arg Phe Ile Lys	ccg ggc 3067 Pro Gly 3255	70
gcc tgg caa gga Ala Trp Gln Gly 3260	Thr Tyr Ala			Leu Met	18
	a Gln Met Glu		tta aga cgc gat Leu Arg Arg Asp 3285		56
			gcc gaa att tat Ala Glu Ile Tyr		14

- 46 -

3290		3295	330	00	
tta ccg Leu Pro 3305	caa gat aaa Gln Asp Lys	a ggc cca ttc s Gly Pro Phe 3310	tcc ctg acg ca Ser Leu Thr Gl 3315	ua gaa atc gag aag n Glu Ile Glu Lys 3320	3
ctg gtg Leu Val	aat gca ggt Asn Ala Gly 3325	y Ser Gly Ser	gcc ggc agt gg Ala Gly Ser Gl 3330	gt aat aat aat ttg Ly Asn Asn Asn Leu 3335	30910 1
gca ttt Ala Phe	ggc gcc ggc Gly Ala Gly 3340	y Thr Asp Thr	aaa act tct tt Lys Thr Ser Le 345	ng cag gca tcc att eu Gln Ala Ser Ile 3350	30 95 8
tca tta Ser Leu	gct gat tta Ala Asp Le 3355	a aaa att cgt u Lys Ile Arg 3360	gag gat tac co Glu Asp Tyr Pr	og gaa tot att ggo ro Glu Ser Ile Gly 3365	31006 Y
aaa ato Lys Ile 3370	Arg Arg Il	c aaa cag atc e Lys Gln Ile 3375	agc gtt acc ct Ser Val Thr Le 338	tg ccg gcg cta tto eu Pro Ala Leu Leo 80	g 31054 u
gga cct Gly Pro 3385	tat cag ga Tyr Gln As	nt gtg cag gca np Val Gln Ala 3390	ata tta tct ta Ile Leu Ser Ty 3395	ac ggc gat aaa gc yr Gly Asp Lys Al 340	a
gga tta Gly Leu	a gcg aac gg n Ala Asn Gl 340	y Cys Ala Ala.	ctg gcc gtt to Leu Ala Val So 3410	cc cac ggt acg aa er His Gly Thr As 3415	t 31150 n
gac ago Asp Sei	ggt caa tt Gly Gln Ph 3420	ne Gln Leu Asp	ttc aac gat g Phe Asn Asp G 3425	gc aaa ttc ctg co Ny Lys Phe Leu Pr 3430	g 31198 o
ttt gaa Phe Gli	a ggt atc go u Gly Ile Al 3435	cc att gat caa la Ile Asp Gln 3440	Gly Thr Leu T	ica ctg agt ttt co fhr Leu Ser Phe Pr 3445	st 31246 To
aat gc Asn Al 345	a Ser Thr P	ca gcc aaa ggt ro Ala Lys Gly 3455	' Lys Gln Ala T	act atg tta aaa ac Thr Met Leu Lys Th 160	cc 31294 ir
ctg aa Leu As 3465	c gat atc a n Asp Ile I	tt ttg cat att le Leu His Ile 3470	cgc tac acc a Arg Tyr Thr I 3475	att aag taa Ile Lys	31336
ccatcc	caac acagaa	ctaa gacaggcco	cc gaatcggggt (ctggtaagga gtttct	atg 31395 Met
cag aa Gln As 3480	t tca cag a n Ser Gln T	ca ttc agc atc hr Phe Ser Mei 3485	g acc gag ctg t t Thr Glu Leu S 3490	tca tta cct aag gg Ser Leu Pro Lys G 34	ly
ggc gg Gly Gl	y Ala Ile T	acc ggt atg gg Thr Gly Met Gly 100	t gaa gca tta a y Glu Ala Leu 1 3505	acg ccg gcc ggg co Thr Pro Ala Gly P 3510	cg 31491 ro
gat gg Asp Gl	nt atg gca g y Met Ala A 3515	occ tta tcg cto Lla Leu Ser Le	g cca ttg ccc a u Pro Leu Pro : 3520	att tct gcc gga c Ile Ser Ala Gly A 3525	gt 31539 rg
ggt ta Gly Ty	at gcc ccc t vr Ala Pro S 3530	cg ctc acg ct Ser Leu Thr Le 353	u Asn Tyr Asn	agc gga acc ggt a Ser Gly Thr Gly A 3540	ac 31587 sn

agc ccg ttc ggt ctc ggt tgg gac tgt aac gtc atg aca att cgt cgt Ser Pro Phe Gly Leu Gly Trp Asp Cys Asn Val Met Thr Ile Arg Arg 3545 3550 3555	31635
cgc acc agt acc ggc gtg ccg aat tat gat gaa acc gat act ttt ctg Arg Thr Ser Thr Gly Val Pro Asn Tyr Asp Glu Thr Asp Thr Phe Leu 3560 3565 3570 3575	31683
ggg ccg gaa ggt gaa gtg ttg gtc gta gca tta aat gag gca ggt caa Gly Pro Glu Gly Glu Val Leu Val Val Ala Leu Asn Glu Ala Gly Gln 3580 3585 3590	31731
gct gat atc cgc agt gaa tcc tca tta cag ggc atc aat ttg ggg atg Ala Asp Ile Arg Ser Glu Ser Ser Leu Gln Gly Ile Asn Leu Gly Met 3595 3600 3605	31779
acc ttc acc gtt acc ggt tat cgc tcc cgt ttg gaa agc cac ttt agc Thr Phe Thr Val Thr Gly Tyr Arg Ser Arg Leu Glu Ser His Phe Ser 3610 3615 3620	31827
cgg ttg gaa tac tgg caa ccc caa aca aca ggc gca acc gat ttc tgg Arg Leu Glu Tyr Trp Gln Pro Gln Thr Thr Gly Ala Thr Asp Phe Trp 3625 3630 3635	31875
ctg ata tac agc ccc gac gga caa gcc cat tta ctg ggc aaa aat cct Leu Ile Tyr Ser Pro Asp Gly Gln Ala His Leu Leu Gly Lys Asn Pro 3640 3645 3650 3655	31923
caa gca cgc atc agc aat cca cta aat gtt aac caa aca gcg caa tgg Gln Ala Arg Ile Ser Asn Pro Leu Asn Val Asn Gln Thr Ala Gln Trp 3660 3665 3670	31971
cta ttg gaa gcc tcg gta tca tcc cac ggc gag cag att tat tat cag Leu Leu Glu Ala Ser Val Ser Ser His Gly Glu Gln Ile Tyr Tyr Gln 3675 3680 3685	32019
tat cga gcc gaa gat gaa act gat tgc gaa act gac gaa ctc aca gcc Tyr Arg Ala Glu Asp Glu Thr Asp Cys Glu Thr Asp Glu Leu Thr Ala 3690 3695 3700	32067
cac ccg aac aca acc gtc cag cgc tac ctg caa gta gta cat tac ggt His Pro Asn Thr Thr Val Gln Arg Tyr Leu Gln Val Val His Tyr Gly 3705 3710 3715	32115
aat cta acc gcc agc gaa gta ttt ccc acg cta aat gga gat gat cca Asn Leu Thr Ala Ser Glu Val Phe Pro Thr Leu Asn Gly Asp Asp Pro 3720 3735 3736	32163
ctc aaa tct ggc tgg ttg ttc tgt tta gta ttt gat tac ggt gag cgc Leu Lys Ser Gly Trp Leu Phe Cys Leu Val Phe Asp Tyr Gly Glu Arg 3740 3745 3750	32211
aaa aac agc tta tct gaa atg ccg cca ttt aaa gcc aca agt aac tgg Lys Asn Ser Leu Ser Glu Met Pro Pro Phe Lys Ala Thr Ser Asn Trp 3755 3760 3765	32259
ctt tgc cgc aaa gac cgt ttt tcc cgt tat gaa tac ggt ttt gca ttg Leu Cys Arg Lys Asp Arg Phe Ser Arg Tyr Glu Tyr Gly Phe Ala Leu 3770 3775 3780	32307
cgc acc cgg cgc tta tgt cgc caa ata ctg atg ttt cac cgt ctg caa Arg Thr Arg Arg Leu Cys Arg Gln Ile Leu Met Phe His Arg Leu Gln 3785 3790 3795	32355
acc ctg tct ggt cag gca aaa ggc gac gat gaa ccc gca tta gtt tca	32403

Thr Leu Ser Gly Gln Ala Lys Gly Asp Asp Glu Pro Ala Leu Val Ser 3800 3805 3810 3815	
cgt ctg ata ctg gat tat gac gaa aac gcg gtg gtc agt acg ctc gtt Arg Leu Ile Leu Asp Tyr Asp Glu Asn Ala Val Val Ser Thr Leu Val 3820 3825 3830	32451
tet gte ege ega gtg gga eat gag eaa gat gge aca aeg geg gte gee Ser Val Arg Arg Val Gly His Glu Gln Asp Gly Thr Thr Ala Val Ala 3835 3840 3845	32499
ctg ccg cca ttg gaa ctg gct tat cag cct ttt gaa cca gaa caa aaa Leu Pro Pro Leu Glu Leu Ala Tyr Gln Pro Phe Glu Pro Glu Gln Lys 3850 3860	32547
gca ctc tgg cga cca atg gat gta ctg gcg aat ttc aac acc atc caa Ala Leu Trp Arg Pro Met Asp Val Leu Ala Asn Phe Asn Thr Ile Gln 3865 3870 3875	32595
cgc tgg caa ctg ctt gat ctg caa ggc gaa ggc gta ccc ggt att ctg Arg Trp Gln Leu Leu Asp Leu Gln Gly Glu Gly Val Pro Gly Ile Leu 3880 3885 3890 3895	32643
tat cag gat aaa aat ggc tgg tgg tat cga tct gct caa cgt cag aca Tyr Gln Asp Lys Asn Gly Trp Trp Tyr Arg Ser Ala Gln Arg Gln Thr 3900 3905 3910	32691
ggg gaa gag atg aat gcg gtc acc tgg ggc aaa atg caa ctc ctt cct Gly Glu Glu Met Asn Ala Val Thr Trp Gly Lys Met Gln Leu Leu Pro 3915 3920 3925	32739
atc acg ccc gct att cag gat aac gcc tca ctg atg gat att aat ggt Ile Thr Pro Ala Ile Gln Asp Asn Ala Ser Leu Met Asp Ile Asn Gly 3930 3935 3940	32787
gat ggg caa ctg gat tgg gtt atc acc ggt ccg ggg cta agg ggt tat Asp Gly Gln Leu Asp Trp Val Ile Thr Gly Pro Gly Leu Arg Gly Tyr 3945 3950 3955	32835
cac agc cag cat cca gat ggc agt tgg aca cgt ttt acg ccg ttg cac His Ser Gln His Pro Asp Gly Ser Trp Thr Arg Phe Thr Pro Leu His 3960 3965 3970 3975	32883
gcc tta ccg ata gaa tat acc cat ccc cgc gcc caa ctt gcg gat tta Ala Leu Pro Ile Glu Tyr Thr His Pro Arg Ala Gln Leu Ala Asp Leu 3980 3985 3990	32931
atg ggg gcc ggg ctg tcc gat tta gtg ctg att ggt ccc aaa agc gtg Met Gly Ala Gly Leu Ser Asp Leu Val Leu Ile Gly Pro Lys Ser Val 3995 4000 4005	32979
cgt ttg tat gcc aat aac cgt gat ggt ttt acc gaa gga cgg gat gtg Arg Leu Tyr Ala Asn Asn Arg Asp Gly Phe Thr Glu Gly Arg Asp Val 4010 4015 4020	33027
gtg caa tcc ggt ggt atc acc ctg ccg tta ccg ggc gcc gat gcg cgt Val Gln Ser Gly Gly Ile Thr Leu Pro Leu Pro Gly Ala Asp Ala Arg 4025 4030 4035	33075
aag tta gtg gcc ttt agc gac gta ctc ggt tca ggc caa gca cat ttg Lys Leu Val Ala Phe Ser Asp Val Leu Gly Ser Gly Gln Ala His Leu 4040 4045 4050 4055	33123
gtt gaa gtt agt gcg acg aaa gtc acc tgc tgg cca aat ctg gga cat Val Glu Val Ser Ala Thr Lys Val Thr Cys Trp Pro Asn Leu Gly His	33171

4060		4065	4070
ggc cgt ttt ggt cag	cca atc aca ttg	Pro Gly Phe Ser G	aa tee gee 33219
Gly Arg Phe Gly Gln	Pro Ile Thr Leu		In Ser Ala
4075	4080		85
gcc aat ttt aat cct Ala Asn Phe Asn Pro 4090	gat cga gtt cat Asp Arg Val His 4095	ctg gcc gat ctg g Leu Ala Asp Leu A 4100	ac ggt agt 33267 sp Gly Ser
ggt cct gcc gat ctg Gly Pro Ala Asp Leu 4105	att tat gtt cat Ile Tyr Val His 4110	gct gac cat ctg g Ala Asp His Leu A 4115	at att ttc 33315 sp Ile Phe
agc aat gaa agt ggt	aac ggt ttt gca	caa cca ttc aca c	etc cgt ttt 33363
Ser Asn Glu Ser Gly	Asn Gly Phe Ala	Gln Pro Phe Thr L	eu Arg Phe
4120	4125	4130	4135
cct gac ggc ctg cgt	Phe Asp Asp Thr	tgc cag cta caa g	tg gct gat 33411
Pro Asp Gly Leu Arg		Cys Gln Leu Gln V	al Ala Asp
4140		4145	4150
gta cag gga tta ggg	gtt gtc agc ctg	atc ctg agc gta c	ro His Met
Val Gln Gly Leu Gly	Val Val Ser Leu	Ile Leu Ser Val P	
4155	4160	41	
gcg cca cac cat tgg Ala Pro His His Trp 4170	cgc tgc gat ctg Arg Cys Asp Leu 4175	acc aac gcg aaa c Thr Asn Ala Lys P 4180	rcg tgg tta 33507 ro Trp Leu
ctc agt gaa atg aac Leu Ser Glu Met Asn 4185	aac aac atg gga Asn Asn Met Gly 4190	gcc cat cac acc c Ala His His Thr L 4195	tg cat tac 33555 eu His Tyr
cgt agc tcc gtc cag	ttt tgg ctg gat	gaa aaa gcc gca g	cc tta gct 33603
Arg Ser Ser Val Gln	Phe Trp Leu Asp	Glu Lys Ala Ala A	la Leu Ala
4200	4205	4210	4215
acc gga caa aca ccg	Val Cys Tyr Leu	ccc ttc ccg gtc c	at acc ctg 33651
Thr Gly Gln Thr Pro		Pro Phe Pro Val H	lis Thr Leu
4220		4225	4230
tgg caa aca gaa acc	gag gat gaa atc	Ser Gly Asn Lys L	ta gtg acc 33699
Trp Gln Thr Glu Thr	Glu Asp Glu Ile		eu Val Thr
4235	4240		45
act tta cgt tac gct Thr Leu Arg Tyr Ala 4250	cac ggc gcc tgg His Gly Ala Trp 4255	gat gga cgt gag c Asp Gly Arg Glu A 4260	gg gaa ttt 33747 rg Glu Phe
cgc ggc ttt ggc tat Arg Gly Phe Gly Tyr 4265	gtt gag cag aca Val Glu Gln Thr 4270	gac agc cat caa c Asp Ser His Gln L 4275	tg gct caa 33795 eu Ala Gln
ggc aat gcg ccg gaa	cgt aca tca ccg	gca ctt acc aaa a	ac tgg tat 33843
Gly Asn Ala Pro Glu	Arg Thr Ser Pro	Ala Leu Thr Lys A	sn Trp Tyr
4280	4285	4290	4295
gcc acc gga atc cct	Glu Val Asp Asn	acg cta tct gcc g	gg tat tgg 33891
Ala Thr Gly Ile Pro		Thr Leu Ser Ala G	ly Tyr Trp
4300		4305	4310
cgc ggt gat acg cag	gct ttc act ggt	ttt acg cca cac t	he Thr Leu
Arg Gly Asp Thr Gln	Ala Phe Thr Gly	Phe Thr Pro His P	
4315	4320	43	

tgg aaa gag ggc aaa gat gtt cca ctg aca ccg gaa gat gac cac aat Trp Lys Glu Gly Lys Asp Val Pro Leu Thr Pro Glu Asp Asp His Asn 4330 4335 4340	33987
ctg tac tgg tta aac cgg gca cta aaa ggt caa cca ctg cgt agt gaa Leu Tyr Trp Leu Asn Arg Ala Leu Lys Gly Gln Pro Leu Arg Ser Glu 4345 4350 4355	34035
ctc tac ggg cta gat ggc agc gca cag cag aag atc ccc tat aca gtg Leu Tyr Gly Leu Asp Gly Ser Ala Gln Gln Lys Ile Pro Tyr Thr Val 4360 4365 4370 4375	34083
act gaa too ogo oca caa gtg ogo caa tta caa gat aac act acc ott Thr Glu Ser Arg Pro Gln Val Arg Gln Leu Gln Asp Asn Thr Thr Leu 4380 4385 4390	34131
tcc ccg gtg ctc tgg gcc tca gtg gtg gaa agt cgt agt tat cac tat Ser Pro Val Leu Trp Ala Ser Val Val Glu Ser Arg Ser Tyr His Tyr 4395 4400 4405	34179
gaa cgt atc atc agc gat ccc caa tgc aat cag gat atc act ctg tcc Glu Arg Ile Ile Ser Asp Pro Gln Cys Asn Gln Asp Ile Thr Leu Ser 4410 4415 4420	34227
agt gac cta ttc ggg caa ccg ctg aaa cag gtt tca gtg caa tat ccc Ser Asp Leu Phe Gly Gln Pro Leu Lys Gln Val Ser Val Gln Tyr Pro 4425 4430 4435	34275
cgc cgc aat aaa cca aca acc aat ccg tat ccc gat aca cta cca gat Arg Arg Asn Lys Pro Thr Thr Asn Pro Tyr Pro Asp Thr Leu Pro Asp 4440 4445 4450 4450	34323
act ctg ttt gcc agc agt tat gac gac caa caa caa cta ttg cgg tta Thr Leu Phe Ala Ser Ser Tyr Asp Asp Gln Gln Gln Leu Leu Arg Leu 4460 4465 4470	34371
acc tac cag caa tcc agt tgg cat cat cta att gct aat gaa ctc aga Thr Tyr Gln Gln Ser Ser Trp His His Leu Ile Ala Asn Glu Leu Arg 4475 4480 4485	34419
gtg tta gga tta ccg gat ggt aca cgc agt gat gct ttc act tac gat Val Leu Gly Leu Pro Asp Gly Thr Arg Ser Asp Ala Phe Thr Tyr Asp 4490 4495 4500	34467
gct aaa cac gtg cct gtt gat ggt tta aat ctg gaa gct cta tgt gct Ala Lys His Val Pro Val Asp Gly Leu Asn Leu Glu Ala Leu Cys Ala 4505 4510 4515	34515
gaa aat agc ctg att gcc gat gat aaa cct cgc gaa tac ctc aac cag Glu Asn Ser Leu Ile Ala Asp Asp Lys Pro Arg Glu Tyr Leu Asn Gln 4520 4525 4530 4535	34563
caa cga acg ttc tat acc gat ggg aaa acc gat gga aaa aat cca acg Gln Arg Thr Phe Tyr Thr Asp Gly Lys Thr Asp Gly Lys Asn Pro Thr 4540 4545 4550	34611
cca ctg aaa aca ccg aca cga cag gct tta atc gcc ttt acc gaa acg Pro Leu Lys Thr Pro Thr Arg Gln Ala Leu Ile Ala Phe Thr Glu Thr 4555 4560 4565	34659
gcg gta tta acg gaa tct ctg tta tcc gca ttt gat ggc ggt atc acg Ala Val Leu Thr Glu Ser Leu Leu Ser Ala Phe Asp Gly Gly Ile Thr 4570 4575 4580	34707

cca gat gaa tta ccc ggc ctt ctg aca caa gca gga tac caa caa gaa Pro Asp Glu Leu Pro Gly Leu Leu Thr Gln Ala Gly Tyr Gln Gln Glu 4585 4590 4595	34755
cct tat ctg ttc cca ctc agt ggc gaa aac caa gtc tgg gta gca cgc Pro Tyr Leu Phe Pro Leu Ser Gly Glu Asn Gln Val Trp Val Ala Arg 4600 4605 4610 4615	34803
aaa ggc tat acc gat tac gga act gag gta caa ttt tgg cgt cct gtc Lys Gly Tyr Thr Asp Tyr Gly Thr Glu Val Gln Phe Trp Arg Pro Val 4620 4625 4630	34851
gca caa cgt aac acc cag tta acc ggg aaa acg act cta aaa tgg gat Ala Gln Arg Asn Thr Gln Leu Thr Gly Lys Thr Thr Leu Lys Trp Asp 4635 4640 4645	34899
acc cac tac tgt gtc atc act caa acc caa gac gcg gct ggt ttg act Thr His Tyr Cys Val Ile Thr Gln Thr Gln Asp Ala Ala Gly Leu Thr 4650 4655 4660	34947
gtc tca gcc aat tat gac tgg cgt ttt ctc aca cct atg caa ctg act Val Ser Ala Asn Tyr Asp Trp Arg Phe Leu Thr Pro Met Gln Leu Thr 4665 4670 4675	34995
gat atc aac gat aat gtg cat atc ata acc ttg gat gcg cta gga cgc Asp Ile Asn Asp Asn Val His Ile Ile Thr Leu Asp Ala Leu Gly Arg 4680 4685 4690 4695	35043
cct gtc act caa cgt ttc tgg gga atc gaa aat ggt gtg gca aca ggt Pro Val Thr Gln Arg Phe Trp Gly Ile Glu Asn Gly Val Ala Thr Gly 4700 4705 4710	35091
tac tct tca cca gaa gca aaa cca ttc act cca cca gtc gat gtc aat Tyr Ser Ser Pro Glu Ala Lys Pro Phe Thr Pro Pro Val Asp Val Asn 4715 4720 4725	35139
gct gcc att gct ctg acc gga cca ctc cct gtc gcg cag tgt ctg gtc Ala Ala Ile Ala Leu Thr Gly Pro Leu Pro Val Ala Gln Cys Leu Val 4730 4735 4740	35187
tat gcg ccg gac agt tgg atg ccg cta ttc ggt cag gaa acc ttc aac Tyr Ala Pro Asp Ser Trp Met Pro Leu Phe Gly Gln Glu Thr Phe Asn 4745 4750 4755	35235
aca tta acg cag gaa gag caa aag aca ctg cgt gat tta cgg att atc Thr Leu Thr Gln Glu Gln Lys Thr Leu Arg Asp Leu Arg Ile Ile 4760 4765 4770 4775	3 52 83
aca gaa gat tgg cgt att tgc gca ctg gct cgc cgc cgt tgg cta caa Thr Glu Asp Trp Arg Ile Cys Ala Leu Ala Arg Arg Arg Trp Leu Gln 4780 4785 4790	35331
agt caa aaa gcc ggc aca cca ttg gtt aag ctg tta acc aac agc atc Ser Gln Lys Ala Gly Thr Pro Leu Val Lys Leu Leu Thr Asn Ser Ile 4795 4800 4805	35379
ggt tta cct ccc cac aac ctc atg ctg gct acg gac cgt tat gac cgt Gly Leu Pro Pro His Asn Leu Met Leu Ala Thr Asp Arg Tyr Asp Arg 4810 4815 4820	35427
gat tot gaa cag caa att ogt caa caa gto goa tto agt gat ggt ttt Asp Ser Glu Gln Gln Ile Arg Gln Gln Val Ala Phe Ser Asp Gly Phe 4825 4830 4835	35475
ggc cgt ttg ttg caa gcg gct gtg cgg cat gag gca ggc gaa gcc tgg	35523

- 52 -

Gly Arg Leu Leu Gln Ala Ala Val Arg His Glu Ala Gly Glu Ala Trp 4840 4845 4850 4855	
caa cgt aac caa gac ggt tct ctg gtg aca aaa atg gaa gat acc aaa Gln Arg Asn Gln Asp Gly Ser Leu Val Thr Lys Met Glu Asp Thr Lys 4860 4865 4870	35571
acg cgc tgg gcg att acg gga cgc act gaa tat gac aat aag ggg cag Thr Arg Trp Ala Ile Thr Gly Arg Thr Glu Tyr Asp Asn Lys Gly Gln 4875 4880 4885	35619
gcg ata cga act tat cag ccc tat ttc ctc aat gac tgg cga tat gtg Ala Ile Arg Thr Tyr Gln Pro Tyr Phe Leu Asn Asp Trp Arg Tyr Val 4890 4895 4900	35667
agt gat gac agc gcc aga aaa gag gcc tat gcc gat act cat atc tat Ser Asp Asp Ser Ala Arg Lys Glu Ala Tyr Ala Asp Thr His Ile Tyr 4905 4910 4915	35715
gat ccg att ggg cgg gaa atc caa gtt atc acg gca aaa ggc tgg ctg Asp Pro Ile Gly Arg Glu Ile Gln Val Ile Thr Ala Lys Gly Trp Leu 4920 4925 4930 4935	35763
cgg cag aac caa tat ttc ccg tgg ttt acc gtg agt gaa gat gaa aat Arg Gln Asn Gln Tyr Phe Pro Trp Phe Thr Val Ser Glu Asp Glu Asn 4940 4945 4950	35811
gat ttg tcc gct gac gcg ctc gtg taa ttgaatcaag attcgctcgt Asp Leu Ser Ala Asp Ala Leu Val 4955 4960	35858
ttaatgttaa cgagcgaata taatatacct aatagatttc gagttgcagc gcggcggca	a 35918
gtgaacgaat ccccaggagc atagataact atgtgactgg ggtgagtgaa agcagccaa	ac 35978
aaagcagcag cttgaaagat gaagggtata aataagaaac tgcattgtga gttctaaat	a 36038
gagtagcagc atattttatt gccttttatt tcataggtaa taaaattcaa ttgctgtaa	aa 36098
aatetgteat eatgagaaet aaaaataaea aetttetett etgeaagaga aateaataa	at 36158
tcaattaaaa atgttataga atctgaatca agaccatttg ttggctcatc aaaaatat	aa 36218
acatccgcat cggtaataaa agctgatgtc aatagaaatt tctttttat cccaagtg	ac 36278
atatgtccat actcaatacc agaataatta gatataccaa aaccatttaa atagtaat	ct 36338'
aattgatatt ttaaattact tttcctataa cgctgactta aattaatcac atccattc	cc 36398
gtgatgaaat tataaaagtt aacattatcc gatagataaa aaccatgctg ttgcaaat	ta 36458
aatoggotot titotooott tittataaaa tiaacoatto ottititaac ottatita	ca 36518
ccagcaatac ttgaaagaaa agtegtttta eeegeeeeat taaeteeege aataeggt	tt 36578
aatccaaccc gaaaatcaca attgactcct gaaaaaatag tcttaccatt aataacaa	.cc 36638
tctaacccaa taacttcaag cataaataac ccctaaaaat aacgtaaaaa agaaaata	ac 36698
accaacaata ataattttcg tgtattgcgt tctcaacaga gaaatagaag aaacaata	at 36 7 58
agaaagaaaa gcataagata aaaatataat cacaggaaaa gatttaacaa caagaaag	ca 36818
aaaaataaaa aaacaaagca aataaaaaaa caaagaaata ccataattaa aaaagaat	at 36878

WO 99/42589

tttccgcaca gataaaaagt tggacaaata tgaaagataa tttatttcaa tatatgatag 36938 attataaaat aacaacatgc atatatataa aacaacactg gcatatatta atgatatata 36998 atcaqcettq tttgggattt gagaaaagge actttcacat aatagatata aaagcagaac 37058 agataatgcc ataataagca cagacatttt atttttatta aaacaataac gaagattcat 37118 tatataaggc aatgaaaaaa aacctgatga aaataatttt ttatttctat taattatata 37178 acatggtgtg aaattcaaat ataatatcaa tgctactaat ggaataacta atgtaaaaat 37238 caaatcatat aatattccac teetgaatga tgeegeeaga agaaagaaca cagcaacaat 37298 aaaaaaatgc aaaaaactta attcaaataa gcaaaatcca attacagcaa aagaaactat 37358 caaaaaaaac acagatgaaa ggtaatgcaa ataattaaca ttttcgtaaa aaaaacctat 37418 aaagaagaaa ataactateg gaaaagcact ataaataaaa aaaacgatac gactaaaaaa 37478 caacgttttt ttacctacca aagaaacgat gattgaattc tcctttgcag aaggaaaaaa 37538 tgattttatg tagccatctg gaataataat attggaagat aaagttatta aaacctcaaa 37658 gataccactg aactttgccg gaagtaataa aagaaaaagg aatataatga catttttatt 37718 cccagacgca aatttettta teetacettt atattecaag geateagega ttattaaatt 37778 catactgcct ctctaaaacc aaaatctaaa taatgtcctt ggtgaatctt tagggaattt 37838 cgtcctggaa tgcaaatata aatagttact gaaaacaata cattgatttt taattaaata 37898 ctggcgatat gaccttaatg atgctacttt attttccagt attcaattcg 37948

<210> 12

<211> 954

<212> PRT

<213> Photorhabdus luminescens

<400> 12

Met Lys Asn Ile Asp Pro Lys Leu Tyr Gln Lys Thr Pro Val Val Asn 1 5 10

Ile Tyr Asp Asn Arg Gly Leu Thr Ile Arg Asn Ile Asp Phe His Arg 20 25 30

Thr Thr Ala Asn Gly Asp Thr Asp Ile Arg Ile Thr Arg His Gln Tyr 35 40 45

Asp Ser Leu Gly His Leu Ser Gln Ser Thr Asp Pro Arg Leu Tyr Glu 50 60

Ala Lys Gln Lys Ser Asn Phe Leu Trp Gln Tyr Asp Leu Thr Gly Asn 65 70 75 80

Ile Leu Cys Thr Glu Ser Val Asp Ala Gly Arg Thr Val Thr Leu Asn 85 90

Asp Ile Glu Gly Arg Pro Leu Leu Thr Val Thr Ala Thr Gly Val Ile 100 105 110

Gln Thr Arg Gln Tyr Glu Thr Ser Ser Leu Pro Gly Arg Leu Leu Ser

		115					120					125			
Val	Thr 130	Glu	Gln	Ile	Pro	Glu 135	Lys	Thr	Ser	Arg	Ile 140	Thr	Glu	Arg	Leu
Ile 145	Trp	Ala	Gly	Asn	Ser 150	Glu	Ala	Glu	Lys	Asn 155	His	Asn	Leu	Ala	Ser 160
Gln	Cys	Val	Arg	His 165	Tyr	Asp	Thr	Ala	Gly 170	Val	Thr	Arg	Leu	Glu 175	Ser
Leu	Ser	Leu	Thr 180		Thr	Val	Leu	Ser 185	Gln	Ser	Ser	Gln	Leu 190	Leu	Ser
Asp	Thr	Glr 195	Glu	Ala	Ser	Trp	Thr 200	Gly	Asp	Asn	Glu	Thr 205	Val	Trp	Gln
Asn	Met 210		ı Ala	Asp	Asp	Ile 215	Tyr	Thr	Thr	Leu	Ser 220	Ala	Phe	Asp	Ala
Thr 225		Alá	a Leu	Leu	Thr 230		Thr	Asp	Ala	Lys 235		Asn	Ile	Gln	Arg 240
Leu	Thr	Туз	. Asp	Val 245		Gly	Gln	Leu	Asn 250		Ser	Trp	Leu	Thr 255	Leu
Lys	AST	Glı	n Pro 260		Gln	Val	Ile	1le 265		Ser	Leu	Thr	Tyr 270		Ala
Ala	Gly	/ GL 27		s Leu	ı Arg	Glu	Glu 280		Gly	Asn	Gly	Val 285	Ile	Thr	Glu
Tyr	Se:		r Glu	ı Pro	Glu	1 Thi 295		ı Glr	Leu	ıle	9 Gly 300		Lys	Thr	His
Arg 305	_	o Se	r Ası	o Ala	a Lys 310		Leu	ı Glr	n Asp	315	ı Arg	ı Tyı	Glu	Tyr	320
Pro	o Va	l Gl	y Asi	n Va. 32!		e Sei	r Ile	e Arg	330		o Ala	a Glu	ı Ala	Thr 335	Arg
Phe	e Tr	p Hi	.s As 34		n Ly:	s Va	l Ala	a Pro 345		l Asi	n Thi	ту:	r Thi 350		: Asp
Se	r Le	u Ty 35		n Le	u Il	e Se	r Ala 36		r Gly	y Ar	g Gl	и Ме 36	t Ala 5	a Ast	ı Ile
Gl	y Gl 37		n Se	r As	n Gl	n Le 37	u Pro 5	o Se	r Le	u Th	r Le 38	u Pr	o Sei	r Ası	o A sn
As 38	_	r T	/r Th	nr As	n Ty 39		r Ar	g Th	т Ту	r Th 39		r As	p Ar	g Gl	y Gly 400
As	n Le	eu Ti	ır Ly	s Il 40		n Hi	s Se	r Se	r Pr 41		a Th	r Gl	n As	n Ası 41	n Tyr 5
Th	r Ti	ur A		le Tr 20	ır Va	l Se	er As	n Ar 42		r As	n Ar	g Al	a Va 43	1 Le 0	u Ser
Th	ır La		hr G] 35	lu As	sp Pr	co Al	la Gl 44		l As	p Al	a Le	u Ph 44	ne As 15	p Al	a Gly
Gl	-	is G 50	ln As	sn Th	r Le		le Se 55	er Gl	y Gl	n As	n Le 46		n Tr	p As	n Thi

Arg 465	Gly	Glu	Leu	Gln	His 470	Val	Thr	Leu	Val	Lys 475	Arg	Asp	Lys	Gly	Ala 480
Asn	Asp	Asp	Arg	Glu 485	Trp	Tyr	Arg	Tyr	Ser 490	Ser	Asp	Gly	Arg	Arg 495	Ile
Leu	Lys	Ile	Asn 500	Glu	Gln	Gln	Thr	Ser 505	Ser	Asn	Ser	Gln	Thr 510	Gln	Arg
Ile	Thr	Tyr 515	Leu	Pro	Ser	Leu	Glu 520	Leu	Arg	Leu	Thr	Gln 525	Asn	Ser	Thr
Ile	Thr 530	Thr	Glu	Asp	Leu	Gln 535	Val	Ile	Thr	Val	Gly 540	Glu	Ala	Gly	Arg
Ala 545	Gln	Val	Arg	Val	Leu 550	His	Trp	Asp	Ser	Gly 5 5 5	Gln	Pro	Glu	Asp	Ile 560
Asp	Asn	Asn	Gln	Leu 565	Arg	Tyr	Ser	Tyr	Asp 570	Asn	Leu	Ile	Gly	Ser 575	Ser
Gln	Leu	Glu	Leu 580	Asp	Ser	Lys	Gly	Glu 585	Ile	Ile	Ser	Glu	Glu 590	Glu	Tyr
Tyr	Pro	Tyr 595	Gly	Gly	Thr	Ala	Leu 600	Trp	Ala	Thr	Arg	Lys 605	Arg	Thr	Glu
Ala	Ser 610	Tyr	Lys	Thr	Ile	Arg 615	Tyr	Ser	Gly	Lys	Glu 620	Arg	Asp	Ala	Thr
Gly 625	Leu	Tyr	Tyr	Tyr	Gly 630	Tyr	Arg	Tyr	Tyr	Gln 635	Pro	Trp	Val	Gly	Arg 640
Trp	Leu	Ser	Ala	Asp 645	Pro	Ala	Gly	Thr	Val 650	Asp	Gly	Leu	Asn	Leu 655	Tyr
Arg	Met	Val	Arg 660	Asn	Asn	Pro	Val	Thr 665		Leu	Asp	Pro	Asp 670	Gly	Leu
Met	Pro	Thr 675	Ile	Ala	Glu	Arg	11e 680		Ala	Leu	Gln	Lys 685		Lys	Val
Ala	Asp 690		Ala	Pro	Ser	Pro 695		Asn	Ala	Thr	Asn 700		Ala	Ile	Asn
Ile 705	Arg	Pro	Pro	Val	Ala 710		Lys	Pro	Thr	Leu 715		Lys	Ala	Ser	Thr 720
Ser	Ser	Gln	Ser	Thr 725		Tyr	Pro	Ile	Lys 730		Ala	Ser	Ile	Lys 735	
Thr	Thr	Ser	Gly 740		Ser	Ile	Thr	Ala 745		Leu	Ser	Pro	Val 750	_	Asn
Lys	Ser	Thr 755	Pro	Glu	Ile	Ser	Leu 760		Glu	Ser	Thr	Gln 765		Asn	Ser
Ser	Ser 770		Ile	Ser	Thr	Asn 775		Gln	Lys	Lys	Ser 780		Thr	Leu	Tyr
Arg 785		Asp) Asn	Arg	Ser 790		e Glu	. Asp	Met	Gln 795		: Lys	Phe	Pro	Glu 800
Gly	Ph∈	Lys	s Ala	Trp 805		Pro	Leu	a Asp	Thr 810		Met	Ala	Arg	Gln 815	

Ala Ser Val Phe Ile Gly Gln Lys Asp Thr Ser Asn Leu Pro Lys Glu 820 825 830

Thr Val Lys Asn Ile Asn Thr Trp Gly Thr Lys Pro Lys Leu Asn Asp 835 840 845

Leu Ser Thr Tyr Ile Lys Tyr Thr Lys Asp Lys Ser Thr Val Trp Val 850 855 860

Ser Thr Ala Ile Asn Thr Glu Ala Gly Gly Gln Ser Ser Gly Ala Pro 865 870 875 880

Leu His Glu Ile Asn Met Asp Leu Tyr Glu Phe Thr Ile Asp Gly Gln 885 890 895

Lys Leu Asn Pro Leu Pro Arg Gly Arg Ser Lys Asp Arg Val Pro Ser 900 905 910

Leu Leu Leu Asp Thr Pro Glu Ile Glu Thr Ala Ser Ile Ile Ala Leu 915 920 925

Asn His Gly Pro Val Asn Asp Ala Glu Val Ser Phe Leu Thr Thr Ile 930 935 940

Pro Leu Lys Asn Val Lys Pro Tyr Lys Arg 945 950

<210> 13

<211> 2522

<212> PRT

<213> Photorhabdus luminescens

<400> 13

Met Ile Leu Lys Gly Ile Asn Met Asn Ser Pro Val Lys Glu Ile Pro 1 5 10 15

Asp Val Leu Lys Ile Gln Cys Gly Phe Gln Cys Leu Thr Asp Ile Ser 20 25 30

His Ser Ser Phe Asn Glu Phe His Gln Gln Val Ser Glu His Leu Ser 35 40 45

Trp Ser Glu Ala His Asp Leu Tyr His Asp Ala Gln Gln Ala Gln Lys 50 60

Asp Asn Arg Leu Tyr Glu Ala Arg Ile Leu Lys Arg Thr Asn Pro Gln 65 70 75 80

Leu Gln Asn Ala Val His Leu Ala Ile Val Ala Pro Asn Ala Glu Leu 85 90 95

Ile Gly Tyr Asn Asn Gln Phe Ser Gly Arg Ala Ser Gln Tyr Val Ala 100 105 110

Pro Gly Thr Val Ser Ser Met Phe Ser Pro Ala Ala Tyr Leu Thr Glu 115 120 125

Leu Tyr Arg Glu Ala Arg Asn Leu His Ala Ser Asp Ser Val Tyr Arg 130 135 140

Leu Asp Thr Arg Arg Pro Asp Leu Lys Ser Met Ala Leu Ser Gln Gln 145 150 155 160

Asn	Met	Asp	Thr	Glu 165	Leu	Ser	Thr	Leu	Ser 170	Leu	Ser	Asn	Glu	Leu 175	Leu
Leu	Glu	Ser	Ile 180	Lys	Thr	Glu	Ser	Lys 185	Leu	Asp	Asn	Tyr	Thr 190	Gln	Val
Met	Glu	Met 195	Leu	Ser	Ala	Phe	Arg 200	Pro	Ser	Gly	Ala	Thr 205	Pro	Tyr	His
Asp	Ala 210	Tyr	Glu	Asn	Val	Arg 215	Lys	Val	Ile	Gln	Leu 220	Gln	Asp	Pro	Gly
Leu 225	Glu	Gln	Leu	Asn	Ala 230	Ser	Pro	Ala	Ile	Ala 235	Gly	Leu	Met	His	Gln 240
Ala	Ser	Leu	Leu	Gly 2 4 5	Ile	Asn	Ala	Ser	Ile 250	Ser	Pro	Glu	Leu	Phe 255	Asn
Ile	Leu	Thr	Glu 260	Glu	Ile	Thr	Glu	Gly 265	Asn	Ala	Glu	Glu	Leu 270	Tyr	Lys
Lys	Asn	Phe 275	Gly	Asn	Ile	Glu	Pro 280	Ala	Ser	Leu	Ala	Met 285	Pro	Glu	Tyr
Leu	Arg 290	Arg	Tyr	Tyr	Asn	Leu 295	Ser	Asp	Glu	Glu	Leu 300	Ser	Gln	Phe	Ile
Gly 305	_	Ala	Ser	Asn	Phe 310	Gly	Gln	Gln	Glu	Туг 315	Ser	Asn	Asn	Gln	Leu 320
Ile	Thr	Pro	Ile	Val 325	Asn	Ser	Asn	Asp	Gly 330		Val	Lys	Val	Тут 335	Arg
Ile	Thr	Arg	Glu 340	_	Thr	Thr	Asn	Ala 345		Gln	Val	Asp	Val 350		Leu
Phe	Pro	Tyr 355	Gly	Gly	Glu	. Asn	Туг 360		Leu	Asn	Tyr	Lys 365		Lys	Asp
Ser	Arg 370		Asp	Val	Ser	Тут 375		Ser	Ile	Lys	Leu 380		Asp	Lys	Arg
Glu 385		Ile	Arg	Ile	Glu 390		Ala	Pro	Glr	Val 395		Il∈	Glu	Tyr	Ser 400
Glu	His	Ile	Thr	Leu 405		Thr	Thr	· Asp	1l∈ 410		Gln	Pro	Phe	Glu 415	
Gly	Leu	Thr	420		Tyr	Pro	Ser	Ser 425		Trp	Ala	Tyr	Ala 430		Ala
Lys	Phe	435	Ile	Glu	Glu	тут	440		Тут	Ser	Phe	Leu 445		Lys	Leu
Asn	1 Lys 450		ı Ile	e Arc	Leu	Ser 455		, Ala	Thr	Glu	1 Let 460		Pro	Thr	·Ile
Leu 465		ı Ser	: Ile	e Val	Arc 470		· Val	Asr	ı Glr	1 Glr 475		ı Asp	Il∈	Asn	Ala 480
Glu	ı Val	Let	ı Gly	485		l Phe	e Leu	ı Thı	Lys 490		тут	. Met	Glr	495	
Alā	a Ile	e Ast	n Ala 500		ı Thi	Ala	a Leu	1 11e 509		ı Cys	s Asr	n Ala	Le. 510		e Ser

Gln Arg Ser Tyr Asp Asn Gln Pro Ser Gln Phe Asp Arg Leu Phe Asn 520 Thr Pro Leu Leu Asn Gly Gln Tyr Phe Ser Thr Gly Asp Glu Glu Ile Asp Leu Asn Pro Gly Ser Thr Gly Asp Trp Arg Lys Ser Val Leu Lys 555 Arg Ala Phe Asn Ile Asp Asp Ile Ser Leu Tyr Arg Leu Lys Ile 565 570 575 Thr Asn His Asn Asn Gln Asp Gly Lys Ile Lys Asn Asn Leu Asn Asn Leu Ser Asp Leu Tyr Ile Gly Lys Leu Leu Ala Glu Ile His Gln Leu Thr Ile Asp Glu Leu Asp Leu Leu Leu Val Ala Val Gly Glu Gly Glu Thr Asn Leu Ser Ala Ile Ser Asp Lys Gln Leu Ala Ala Leu Ile Arg Lys Leu Asn Thr Ile Thr Val Trp Leu Gln Thr Gln Lys Trp Ser Ala 650 Phe Gln Leu Phe Val Met Thr Ser Thr Ser Tyr Asn Lys Thr Leu Thr 665 Pro Glu Ile Lys Asn Leu Leu Asp Thr Val Tyr His Gly Leu Gln Gly Phe Asp Lys Asp Lys Ala Asn Leu Leu His Val Met Ala Pro Tyr Ile Ala Ala Thr Leu Gln Leu Ser Ser Glu Asn Val Ala His Ser Val Leu Leu Trp Ala Asp Lys Leu Lys Pro Gly Asp Gly Ala Met Thr Ala Glu 725 730 735 Lys Phe Trp Asp Trp Leu Asn Thr Gln Tyr Thr Pro Asp Ser Ser Glu Val Leu Ala Thr Gln Glu His Ile Val Gln Tyr Cys Gln Ala Leu Ala Gln Leu Glu Met Val Tyr His Ser Thr Gly Ile Asn Glu Asn Ala Phe Arg Leu Phe Val Thr Lys Pro Glu Met Phe Gly Ser Ser Thr Glu Ala 795 Val Pro Ala His Asp Ala Leu Ser Leu Ile Met Leu Thr Arg Phe Ala Asp Trp Val Asn Ala Leu Gly Glu Lys Ala Ser Ser Val Leu Ala Ala Phe Glu Ala Asn Ser Leu Thr Ala Glu Gln Leu Ala Asp Ala Met Asn

Leu Asp Ala Asn Leu Leu Gln Ala Ser Thr Gln Ala Gln Asn His

	850					855					860				
Gln 865	His	Leu	Pro	Pro	Val 870	Thr	Gln	Lys	Asn	Ala 875	Phe	Ser	Cys	Trp	Thr 880
Ser	Ile	Asp	Thr	Ile 885	Leu	Gln	Trp	Val	Asn 890	Val	Ala	Gln	Gln	Leu 895	Asn
Val	Ala	Pro	Gln 900	Gly	Val	Ser	Ala	Leu 905	Val	Gly	Leu	Asp	Tyr 910	Ile	Gln
Leu	Asn	Gln 915	Lys	Ile	Pro	Thr	Тут 920	Ala	Gln	Trp	Glu	Ser 925	Ala	Gly	Glu
Ile	Leu 930	Thr	Ala	Gly	Leu	Asn 935	Ser	Gln	Gln	Ala	Asp 940	Ile	Leu	His	Ala
Phe 945	Leu	Asp	Glu	Ser	Arg 950	Ser	Ala	Ala	Leu	Ser 955	Thr	Tyr	Tyr	Ile	Arg 960
Gln	Val	Ala	Lys	Pro 965	Ala	Ala	Ala	Ile	Lys 970	Ser	Arg	Asp	As p	Leu 975	Tyr
Gln	Tyr	Leu	Leu 980	Ile	Asp	Asn	Gln	Val 985	Ser	Ala	Ala	Ile	Lys 990	Thr	Thr
Arg	Ile	Ala 995	Glu	Ala	Ile		Ser 1000	Ile	Gln	Leu	_	Val 1005		Arg	Thr
	Glu 1010		Val	Glu		Asn 1015		His	Ser	_	Val 1020		Ser	Arg	Gln
025			Asp		1030	_				1035	_				1040
				1045					105 0					1055	
	_		: Gly 1060			_		1065	,				1070		
		1075					1080	ı				1085	5		
-	1090	1	Ser			1095	,				1100)			
105			Asn		1110)				1115	5				1120
			ı Thr	1125	,				1130)				1135	.
			Ser 1140)				1145	5				1150)	
	-	1155					1160)				1169	5		
	1170)	c Lys			1179	5			•	1180)		-	
11e 185		. Lys	s Glr	נולו ר	1190		n Sei	: Ly:	s Asq	o Gly 1199		r Gli	n Thu	Glu	1200

- Asp Tyr Arg Tyr Glu Leu Lys Leu Ala His Ile Arg Tyr Asp Gly Thr 1205 1210 1215
- Trp Asn Thr Pro Ile Thr Phe Asp Val Asn Glu Lys Ile Ser Lys Leu 1220 1225 1230
- Glu Leu Ala Lys Asn Lys Ala Pro Gly Leu Tyr Cys Ala Gly Tyr Gln 1235 1240 1245
- Gly Glu Asp Thr Leu Leu Val Met Phe Tyr Asn Gln Gln Asp Thr Leu 1250 1255 1260
- Asp Ser Tyr Lys Thr Ala Ser Met Gln Gly Leu Tyr Ile Phe Ala Asp 265 1270 1275 1280
- Met Glu Tyr Lys Asp Met Thr Asp Gly Gln Tyr Lys Ser Tyr Arg Asp 1285 1290 1295
- Asn Ser Tyr Lys Gln Phe Asp Thr Asn Ser Val Arg Arg Val Asn Asn 1300 1305 1310
- Arg Tyr Ala Glu Asp Tyr Glu Ile Pro Ser Ser Val Asn Ser Arg Lys 1315 1320 1325
- Gly Tyr Asp Trp Gly Asp Tyr Tyr Leu Ser Met Val Tyr Asn Gly Asp 1330 1335 1340
- Ile Pro Thr Ile Ser Tyr Lys Ala Thr Ser Ser Asp Leu Lys Ile Tyr 345 1350 1355 1360
- Ile Ser Pro Lys Leu Arg Ile Ile His Asn Gly Tyr Glu Gly Gln Gln 1365 1370 1375
- Arg Asn Gln Cys Asn Leu Met Asn Lys Tyr Gly Lys Leu Gly Asp Lys 1380 1385 1390
- Phe Ile Val Tyr Thr Ser Leu Gly Val Asn Pro Asn Asn Ser Ser Asn 1395 1400 1405
- Lys Leu Met Phe Tyr Pro Val Tyr Gln Tyr Asn Gly Asn Val Ser Gly 1410 1415 1420
- Leu Ser Gln Gly Arg Leu Leu Phe His Arg Asp Thr Asn Tyr Ser Ser 425 1430 1435 1440
- Lys Val Glu Ala Trp Ile Pro Gly Ala Gly Arg Ser Leu Thr Asn Pro 1445 1450 1455
- Asn Ala Ala Ile Gly Asp Asp Tyr Ala Thr Asp Ser Leu Asn Lys Pro 1460 1465 1470
- Asn Asp Leu Lys Gln Tyr Val Tyr Met Thr Asp Ser Lys Gly Thr Ala 1475 1480 1485
- Thr Asp Val Ser Gly Pro Val Asp Ile Asn Thr Ala Ile Ser Pro Ala 1490 1495 1500
- Lys Val Gln Val Thr Val Lys Ala Gly Ser Lys Glu Gln Thr Phe Thr 505 1510 1515 1520
- Ala Asp Lys Asn Val Ser Ile Gln Pro Ser Pro Ser Phe Asp Glu Met 1525 1530 1535
- Asn Tyr Gln Phe Asn Ala Leu Glu Ile Asp Gly Ser Ser Leu Asn Phe 1540 1545 1550

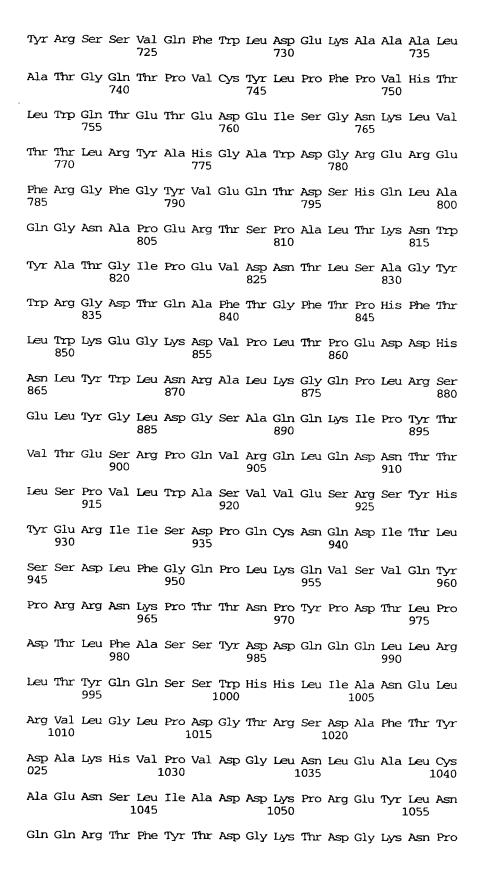
- Thr Asn Asn Ser Ala Ser Ile Asp Ile Thr Phe Thr Ala Phe Ala Glu 1555 1560 1565
- Asp Gly Arg Lys Leu Gly Tyr Glu Ser Phe Ser Ile Pro Ile Thr Arg 1570 1575 1580
- Lys Val Ser Thr Asp Asn Ser Leu Thr Leu Arg His Asn Glu Asn Gly 585 1590 1595 1600
- Ala Gln Tyr Met Gln Trp Gly Val Tyr Arg Ile Arg Leu Asn Thr Leu 1605 1610 1615
- Phe Ala Arg Gln Leu Val Ala Arg Ala Thr Thr Gly Ile Asp Thr Ile 1620 1625 1630
- Leu Ser Met Glu Thr Gln Asn Ile Gln Glu Pro Gln Leu Gly Lys Gly 1635 1640 1645
- Phe Tyr Ala Thr Phe Val Ile Pro Pro Tyr Asn Pro Ser Thr His Gly 1650 1655 1660
- Asp Glu Arg Trp Phe Lys Leu Tyr Ile Lys His Val Val Asp Asn Asn 665 1670 1675 1680
- Ser His Ile Ile Tyr Ser Gly Gln Leu Lys Asp Thr Asn Ile Ser Thr 1685 1690 1695
- Thr Leu Phe Ile Pro Leu Asp Asp Val Pro Leu Asn Gln Asp Tyr Ser 1700 1705 1710
- Ala Lys Val Tyr Met Thr Phe Lys Lys Ser Pro Ser Asp Gly Thr Trp 1715 1720 1725
- Trp Gly Pro His Phe Val Arg Asp Asp Lys Gly Ile Val Thr Ile Asn 1730 1735 1740
- Pro Lys Ser Ile Leu Thr His Phe Glu Ser Val Asn Val Leu Asn Asn 745 1750 1755 1760
- Ile Ser Ser Glu Pro Met Asp Phe Ser Gly Ala Asn Ser Leu Tyr Phe 1765 1770 1775
- Trp Glu Leu Phe Tyr Tyr Thr Pro Met Leu Val Ala Gln Arg Leu Leu 1780 1785 1790
- His Glu Gln Asn Phe Asp Glu Ala Asn Arg Trp Leu Lys Tyr Val Trp 1795 1800 1805
- Ser Pro Ser Gly Tyr Ile Val His Gly Gln Ile Gln Asn Tyr Gln Trp 1810 1815 1820
- Asn Val Arg Pro Leu Leu Glu Asp Thr Ser Trp Asn Ser Asp Pro Leu 825 1830 1835 1840
- Asp Ser Val Asp Pro Asp Ala Val Ala Gln His Asp Pro Met His Tyr 1845 1850 1855
- Lys Val Ser Thr Phe Met Arg Thr Leu Asp Leu Leu Ile Ala Arg Gly 1860 1865 1870
- Asp His Ala Tyr Arg Gln Leu Glu Arg Asp Thr Leu Asn Glu Ala Lys 1875 1880 1885
- Met Trp Tyr Met Gln Ala Leu His Leu Leu Gly Asp Lys Pro Tyr Leu

1	890				18	3 9 5				1	900				
Pro 905	Leu	Ser '	Thr	Thr '	Trp 2 910	Asn <i>l</i>	Asp B	Pro i	Arg 1	Leu . 915	Asp 1	Lys A	Ala A	Ala 1	Asp 920
Ile	Thr	Thr		Ser . .925	Ala 1	His S	Ser S	Ser :	Ser 930	Ile	Val 1	Ala I	ieu i	Arg (935	Gln
Ser	Thr		Ala 940	Leu	Leu	Ser 1	Leu A	Arg 945	Ser	Ala	Asn '	Thr 1	Leu ' 950	Thr .	Asp
Leu		Leu 1955	Pro	Gln	Ile .		Glu \ 960	<i>V</i> al	Met	Met	Asn 1	Tyr ' 965	Trp (Gln	Thr
	Ala 1970	Gln	Arg	Val		Asn 975	Leu i	Arg	His	Asn 1	Leu 980	Ser	Ile.	Asp	Gly
Gln 985	Pro	Leu	Tyr	Leu 1	Pro 1990	Ile	Tyr .	Ala		Pro 1995	Ala	Asp	Pro	Lys 2	Ala 2000
Leu	Leu	Ser		Ala 2005	Val	Ala	Thr		Gln :010	Gly	Gly	Gly	Lys 2	Leu 2015	Pro
Glu	Ser		Met 2020	Ser	Leu	Trp		Phe :025	Pro	His	Met	Leu 2	Glu !030	Asn	Ala
Arg	Ser	Met 2035	Val	Ser	Gln		Thr 2040	Gln	Phe	Gly		Thr 2045	Leu	Gln	Asn
	Il∈ 2050		Arg	Gln		Ala 2 05 5	Glu	Ala	Leu		Ala 2060	Leu	Leu	Gln	Asn
Gln 065		Ala	Glu	Leu	11e 2070	Leu	Thr	Asn		Ser 20 7 5		Gln	Asp	Lys	Thr 2080
Il∈	: Glu	ı Glu	Leu	Asp 2085		Glu	Lys	Thr	Val 2090	Leu	Glu	Lys	Ser	Lys 2095	Ala
Gly	Alá	a Glr	Sei 2100	Arg	Phe	Asp		Tyr 2105		Lys	Leu		Asp 2110		Asn
Ile	e Asi	n Ala 2115		y Glu	ı Asn		Ala 2120		Thr	Leu	a Arg	Ala 2125		Ala	Ala
Gly	213		Th	r Ala	a Val	Gln 2135		Ser	Arg	j Lei	1 Ala 2140		Ala	Ala	Ala
Ası 149	_	u Va	l Pro	o Ast	11e 2150		: Gly	Phe	Ala	a Gly 215		, Gly	Ser	· Arg	Trp 2160
Gl	y Al	a Ile	e Al	a Glu 216		a Thr	Gly	туг	Va. 21 7 0		t Glu	ı Ph∈	e Ser	Alá 2175	a Asn
Va	l Me	t Ası	n Th 218		u Ala	a Asp	Lys	2185		r Gli	n Sei	c Glu	1 Thr 2190	: Tyı)	Arg
Ar	g Ar	g Ar 219		n Gl	u Tr	o Glu	1 lle 2200		n Ar	g As	n Ası	n Ala 2205	a Glu	ı Ala	a Glu
Le	u Ly 221		n Le	eu As	p Ala	a Glr 221		ı Ly:	s Se	r Le	u Ala 222		l Arg	g Ar	g Glu
Al 22	_	a Va	l Le	eu Gl	n Ly 223		r Sei	Le	u Ly	rs Th 223		n Gl	n Gli	u Gli	n Thr 2240

- Gln Ala Gln Leu Ala Phe Leu Gln Arg Lys Phe Ser Asn Gln Ala Leu 2245 2250 2255
- Tyr Asn Trp Leu Arg Gly Arg Leu Ala Ala Ile Tyr Phe Gln Phe Tyr 2260 2265 2270
- Asp Leu Ala Ile Ala Arg Cys Leu Met Ala Glu Gln Ala Tyr Arg Trp 2275 2280 2285
- Glu Ile Ser Asp Asp Ser Ala Arg Phe Ile Lys Pro Gly Ala Trp Gln 2290 2295 2300
- Gly Thr Tyr Ala Gly Leu Leu Ala Gly Glu Thr Leu Met Leu Ser Leu 305 2310 2315 2320
- Ala Gln Met Glu Asp Ala His Leu Arg Arg Asp Lys Arg Ala Leu Glu 2325 2330 2335
- Val Glu Arg Thr Val Ser Leu Ala Glu Ile Tyr Ala Gly Leu Pro Gln 2340 2345 2350
- Asp Lys Gly Pro Phe Ser Leu Thr Gln Glu Ile Glu Lys Leu Val Asn 2355 2360 2365
- Ala Gly Ser Gly Ser Ala Gly Ser Gly Asn Asn Leu Ala Phe Gly 2370 2375 2380
- Ala Gly Thr Asp Thr Lys Thr Ser Leu Gln Ala Ser Ile Ser Leu Ala 385 2390 2395 2400
- Asp Leu Lys Ile Arg Glu Asp Tyr Pro Glu Ser Ile Gly Lys Ile Arg 2405 2410 2415
- Arg Ile Lys Gln Ile Ser Val Thr Leu Pro Ala Leu Leu Gly Pro Tyr 2420 2425 2430
- Gln Asp Val Gln Ala Ile Leu Ser Tyr Gly Asp Lys Ala Gly Leu Ala 2435 2440 2445
- Asn Gly Cys Ala Ala Leu Ala Val Ser His Gly Thr Asn Asp Ser Gly 2450 2455 2460
- Gln Phe Gln Leu Asp Phe Asn Asp Gly Lys Phe Leu Pro Phe Glu Gly 465 2470 2475 2480
- Ile Ala Ile Asp Gln Gly Thr Leu Thr Leu Ser Phe Pro Asn Ala Ser 2485 2490 2495
- Thr Pro Ala Lys Gly Lys Gln Ala Thr Met Leu Lys Thr Leu Asn Asp 2500 2505 2510
- Ile Ile Leu His Ile Arg Tyr Thr Ile Lys
 2515 2520
- <210> 14
- <211> 1481
- <212> PRT
- <213> Photorhabdus luminescens
- <400> 14
- Met Gln Asn Ser Gln Thr Phe Ser Met Thr Glu Leu Ser Leu Pro Lys 1 5 10 15
- Gly Gly Gly Ala Ile Thr Gly Met Gly Glu Ala Leu Thr Pro Ala Gly

			20					25					30		
Pro	Asp	Gly 35	Met	Ala	Ala	Leu	Ser 40	Leu	Pro	Leu	Pro	Ile 45	Ser	Ala	Gly
Arg	Gly 50	Tyr	Ala	Pro	Ser	Leu 55	Thr	Leu	Asn	Tyr	Asn 60	Ser	Gly	Thr	Gly
Asn 65	Ser	Pro	Phe	Gly	Leu 70	Gly	Trp	Asp	Cys	Asn 75	Val	Met	Thr	Ile	Arg 80
Arg	Arg	Thr	Ser	Thr 85	Gly	Val	Pro	Asn	Tyr 90	Asp	Glu	Thr	Asp	Thr 95	Phe
Leu	Gly	Pro	Glu 100	Gly	Glu	Val	Leu	Val 105	Val	Ala	Leu	Asn	Glu 110	Ala	Gly
Gln	Ala	Asp 115	Ile	Arg	Ser	Glu	Ser 120	Ser	Leu	Gln	Gly	Ile 125	Asn	Leu	Gly
Met	Thr 130	Phe	Thr	Val	Thr	Gly 135	Tyr	Arg	Ser	Arg	Leu 140	Glu	Ser	His	Phe
Ser 145	Arg	Leu	Glu	Tyr	Trp 150	Gln	Pro	Gln	Thr	Thr 155	Gly	Ala	Thr	Asp	Phe 160
Trp	Leu	Ile	Tyr	Ser 165	Pro	Asp	Gly	Gln	Ala 170	His	Leu	Leu	Gly	Lys 175	Asn
Pro	Gln	Ala	Arg 180	Ile	Ser	Asn	Pro	Leu 185	Asn	Val	Asn	Gln	Thr 190	Ala	Gln
Trp	Leu	Leu 195	Glu	Ala	Ser	Val	Ser 200	Ser	His	Gly	Glu	Gln 205	Ile	Tyr	Tyr
Gln	Туг 210	Arg	Ala	Glu	Asp	Glu 215	Thr	Asp	Cys	Glu	Thr 220	Asp	Glu	Leu	Thr
Ala 225	His	Pro	Asn	Thr	Thr 230	Val	Gln	Arg	Tyr	Leu 235	Gln	Val	Val	His	Tyr 240
Gly	Asn	Leu	Thr	Ala 245	Ser	Glu	Val	Phe	Pro 250		Leu	Asn	Gly	Asp 255	Asp
Pro	Leu	Lys	Ser 260	Gly	Trp	Leu	Phe	Cys 265	Leu	Val	Phe	Asp	Туr 270	Gly	Glu
Arg	Lys	Asn 275		Leu	Ser	Glu	Met 280		Pro	Phe	Lys	Ala 285		Ser	Asn
Trp	Leu 290		Arg	Lys	Asp	Arg 295	Phe	Ser	Arg	Tyr	Glu 300	_	Gly	Phe	Ala
Leu 305		Thr	Arg	Arg	Leu 310		Arg	Gln	Ile	Leu 315		Phe	His	Arg	Leu 320
Gln	Thr	Leu	Ser	Gly 325		Ala	Lys	Gly	Asp 330		Glu	Pro	Ala	Leu 335	
Ser	Arg	Leu	11e 340		Asp	Tyr	Asp	Glu 345		Ala	Val	Val	Ser 350		Leu
Val	Ser	Val 355		Arg	Val	Gly	His 360		Gln	Asp	Gly	Thr 365		Ala	Val

Ala	Leu 370	Pro	Pro	Leu	Glu	Leu 375	Ala	Tyr	Gln	Pro	Phe 380	Glu	Pro	Glu	Gln
Lys 385	Ala	Leu	Trp	Arg	Pro 390	Met	Asp	Val	Leu	Ala 395	Asn	Phe	Asn	Thr	Ile 400
Gln	Arg	Trp	Gln	Leu 405	Leu	Asp	Leu	Gln	Gly 410	Glu	Gly	Val	Pro	Gly 41 5	Ile
Leu	Tyr	Gln	Asp 420	Lys	Asn	Gly	Trp	Trp 425	Tyr	Arg	Ser	Ala	Gln 430	Arg	Gln
Thr	Gly	Glu 435	Glu	Met	Asn	Ala	Val 440	Thr	Trp	Gly	Lys	Met 445	Gln	Leu	Leu
Pro	Ile 4 50	Thr	Pro	Ala	Ile	Gln 455	Asp	Asn	Ala	Ser	Leu 460	Met	Asp	Ile	Asn
Gly 465	Asp	Gly	Gln	Leu	Asp 470	Trp	Val	Ile	Thr	Gly 475	Pro	Gly	Leu	Arg	Gly 480
Tyr	His	Ser	Gln	His 485	Pro	Asp	Gly	Ser	Trp 490	Thr	Arg	Phe	Thr	Pro 495	Leu
His	Ala	Leu	Pro 500	Ile	Glu	Tyr	Thr	His 505		Arg	Ala	Gln	Leu 510	Ala	Asp
Leu	Met	Gly 515	Ala	Gly	Leu	Ser	Asp 520	Leu	Val	Leu	Ile	Gly 525	Pro	Lys	Ser
Val	Arg 530		Tyr	Ala	Asn	Asn 535	_	Asp	Gly	Phe	Thr 540	Glu	Gly	Arg	Asp
Val 545		Gln	Ser	Gly	Gly 550	Ile	Thr	Leu	Pro	Leu 555		Gly	Ala	Asp	Ala 560
Arg	Lys	Leu	Val	Ala 565		Ser	Asp	Val	Leu 570		Ser	Gly	Gln	Ala 575	
Leu	Val	Glu	Val 580		Ala	Thr	Lys	Val 585		Cys	Trp	Pro	Asn 590		Gly
His	Gly	Arg 595	Phe	Gly	Glr	Pro	000 600		Leu	Pro	Gly	Phe 605		Gln	Ser
Ala	Ala 610		Phe	Asn	Pro	Asp 615	-	(Val	. His	s Leu	Ala 620	_	Leu	Asp	Gly
Ser 625		Pro	Ala	Asp	630		e Tyr	· Val	. His	635		His	Leu	Asp	640
Phe	e Ser	Asr	Glu	Ser 645		/ Asr	ı Gly	/ Phe	e Ala 650		n Pro	Phe	e Thr	655	
Phe	e Pro) Asp	Gly 660		ı Arç	y Phe	e Asp	Asp 669		c Cys	s Glr	ı Let	Glr 670		Ala
Asţ	o Val	Glr 675	n Gly	/ Leu	ı Gly	/ Val	l Val 680		Le	ı Ile	e Leu	Ser 685		Pro	His
Met	Ala 690		His	s His	Tr	Arg 699		a Asp	p Lei	ı Thi	r Ası 700		a Lys	s Pro	Trp
Let 705		ו Se	c Glu	א Met	Ası 710		n Asn	n Met	c Gly	y Ala 71		s His	s Thi	Lev	His 720



1065 1070 1060 Thr Pro Leu Lys Thr Pro Thr Arg Gln Ala Leu Ile Ala Phe Thr Glu 1080 Thr Ala Val Leu Thr Glu Ser Leu Leu Ser Ala Phe Asp Gly Gly Ile 1095 1100 Thr Pro Asp Glu Leu Pro Gly Leu Leu Thr Gln Ala Gly Tyr Gln Gln 1110 1115 Glu Pro Tyr Leu Phe Pro Leu Ser Gly Glu Asn Gln Val Trp Val Ala 1130 Arg Lys Gly Tyr Thr Asp Tyr Gly Thr Glu Val Gln Phe Trp Arg Pro 1145 Val Ala Gln Arg Asn Thr Gln Leu Thr Gly Lys Thr Thr Leu Lys Trp 1160 Asp Thr His Tyr Cys Val Ile Thr Gln Thr Gln Asp Ala Ala Gly Leu 1175 Thr Val Ser Ala Asn Tyr Asp Trp Arg Phe Leu Thr Pro Met Gln Leu 1190 Thr Asp Ile Asn Asp Asn Val His Ile Ile Thr Leu Asp Ala Leu Gly 1205 1210 Arg Pro Val Thr Gln Arg Phe Trp Gly Ile Glu Asn Gly Val Ala Thr Gly Tyr Ser Ser Pro Glu Ala Lys Pro Phe Thr Pro Pro Val Asp Val Asn Ala Ala Ile Ala Leu Thr Gly Pro Leu Pro Val Ala Gln Cys Leu 1255 Val Tyr Ala Pro Asp Ser Trp Met Pro Leu Phe Gly Gln Glu Thr Phe 1270 1275 Asn Thr Leu Thr Gln Glu Glu Gln Lys Thr Leu Arg Asp Leu Arg Ile Ile Thr Glu Asp Trp Arg Ile Cys Ala Leu Ala Arg Arg Arg Trp Leu Gln Ser Gln Lys Ala Gly Thr Pro Leu Val Lys Leu Leu Thr Asn Ser 1320 Ile Gly Leu Pro Pro His Asn Leu Met Leu Ala Thr Asp Arg Tyr Asp 1335 Arg Asp Ser Glu Gln Gln Ile Arg Gln Gln Val Ala Phe Ser Asp Gly 1355 1350 Phe Gly Arg Leu Leu Gln Ala Ala Val Arg His Glu Ala Gly Glu Ala 1370

Trp Gln Arg Asn Gln Asp Gly Ser Leu Val Thr Lys Met Glu Asp Thr 1385

Lys Thr Arg Trp Ala Ile Thr Gly Arg Thr Glu Tyr Asp Asn Lys Gly 1400

1405

Gln Ala Ile Arg Thr Tyr Gln Pro Tyr Phe Leu Asn Asp Trp Arg Tyr 1410 1415 1420	
Val Ser Asp Asp Ser Ala Arg Lys Glu Ala Tyr Ala Asp Thr His Ile 425 1430 1435 1440	
Tyr Asp Pro Ile Gly Arg Glu Ile Gln Val Ile Thr Ala Lys Gly Trp 1445 1450 1455	
Leu Arg Gln Asn Gln Tyr Phe Pro Trp Phe Thr Val Ser Glu Asp Glu 1460 1465 1470	
Asn Asp Leu Ser Ala Asp Ala Leu Val 1475 1480	
<210> 15 <211> 23 <212> DNA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonucleotide	
<400> 15 cgggatccga tgattttaaa agg	23
<210> 16 <211> 16 <212> DNA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonucleotide	
<400> 16 gcgccattga tttgag	16
<210> 17 <211> 19 <212> DNA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonucleotide	
<400> 17 cattagaggt cgaacgtac	19
<210> 18 <211> 26 <212> INVA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonucleotide	
<400> 18 gagcgagctc ttacttaatg gtgtag	26
<210> 19	

<211> 28 <212> DNA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonucleotide	
<400> 19 cagegagete catgeagaat teacagae	28
<210> 20 <211> 18 <212> DNA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonucleotide	
<400> 20 ggcaatggca gcgataag	18
<210> 21 <211> 18 <212> DNA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonucleotide	
<400> 21 cattaacgca ggaagagc	18
<210> 22 <211> 26 <212> DNA <213> Artificial Sequence	
<pre><220> <223> Description of Artificial Sequence:oligonucleotide</pre>	
<400> 22 gacctcgagt tacacgagcg cgtcag	26

```
SEQUENCE LISTENS
<110> Novactie AG
<120> Nove), Toxins And Uses Thereof
<130> PI/5-30421/A/CGC 1963
<140>
<141>
<160-22
<170> PatentIn Ver. 2.0
<210> 1
<211> 9717
<212> NA
<213> Photorhabdus luminescens
<220>
<221> CDS
<222> {412}..(1665)
<223> arf1 ~16.4 kbs
<220>
<221> cms
<222> (1686) ... (2447)
<223> orf2 ~28.1kDa
<220×
<221> CDS
<2225 {2750}..(3310)
<223> orf3 ~20.7 kma
<220>
<221> CDS
<222> (3342)..(4118)
<223> ort4 -28.7 kDa
<220≻
<221> COS
<222> (4515) . . (9269)
<223> orf5 -176 kDa.
gwattcatat gotatgaeat asacegitigg cycaataett asacciatis tittiatiit 60
gttüttäten aatgetatgo tittettaasn agaatastga gttästgota aataaatoot 120
oggstitet cetgetetta iggccgesig igstgigesc wattettite taattageti 180
aztaslalma iggintiama etangental atlitaliost gogitattiat catoggitti 240
attacatggy gaalaalola taaattagti tiacataatt cacaaatagc qattacatta 300
accoggaata trasaaatak tratttatga tratggtgat atatetteat tageetaett 360
ttataactag aaaaattgac attttcaate catgtataaa tggtaaccaa t atg cag 417
                                                            Met Gln
```

age get cas cas get get att acs gat etg ggt gen ges acs cog att Arg Ala Oln Arg Val Val Ile Thr Gly Wet Gly Ala Val Thr Pro Ile 5

ggr. Gly	gaa Glu 20	gat Asp	gtt Val	gaa Glu	Ser Ser	tgt Cys 25	tgg Trp	caa Gln	agt Ser	att Ile	att Ilo 30	gaa Glu	Liye Liye	caa Glo	cat. His	513
eça Arg 35	ttt Phie	cac His	aga Arg	att Ile	ന്തും Glu 40	ttt Phe	cet Pro	god Asp	tca Ser	tte Phe 45	att Ile	aat Ayrı	teg Ser	egt. Arg	tta The 50	561
ttt Phe	tet Ser	tto Ehe	ett Leu	gra Ala 55	cca Pro	aac Asn	cica Pro	tice Ser	ege Arg 60	tat Tyr	cag Gln	tta Len	tta føk	oca Pro 65	aaa lys	609
аэд Јус	ttg Jeu	aet Thr	cat Hie 70	aca Thir	ctt Leu	tct Ser	gac Asp	bgc Cys 75	gga Gly	еса Lys	gca Ale	gca Ala	ttg Leu 80	TAS Seà	gng Ala	657
act Thr	tat Tyr	caa Glin 85	get Ala	ttt Phe	ace Thr	caz Gln	908 Ala 90	tte Phe	CJÀ 33c	Asj aça	aat Aen	ata I) e 95	tca Sor	cct Pro	gtt Val	705
gaz Glu	tati Tyr 100	tær: Tyr	gest. Asp	nys aea	tec Tyr	984 Glu 3 0 5	tgt Cys	gjà gàc	gts Val	att Ile	ctt Leu 110	Gly 930	agt Ser	ggt Gly	tgg Trp	753
ggs Gly 115	g¢t. Ala	stt Ilb	get Asp	aat Asn	god Ala 120	eja Gre	get. Asp	ræt His	got Ala	tec Cys 125	caa Gln	tat Tyr	aag Tyrs	caz Gln	gca Ala 130	601
aea Lys	tta Icu	get Ala	cat His	eet Pxo 135	atg Met	agt Sær	est Asn	ett Iæ	att Ile 140	acc Trur	6bg Met	cca Pro	agc Ser	tee Ser 145	atg Met	849
apr. ect	get Ala	gca Ala	tgt Cys 150	tog Ser	att Ile	atg Met	tat Tyr	9923 Gly 155	eta Leu	egt Arg	g gt Gûy	tat Tyr	caa Gla 160	aat Abn	acc Thr	897
gtt Val	ə(g Met	got Ala 165	goc Ala	tge Cys	gca Ala	acg Tiur	990 Gly 170	aca Thr	atg Met	YJ⊊ G©3	ata Ile	990 Oly 175	gst. Asp	gur: Ala	tiut. Phe	945
gaa Glu	att Ile 180	att Ilc	oge Az g	tca Sor	977 953	rgg Arg 185	gça Ala	rys eaa	tgt Cya	atg M et	att Ile 190	YJa gcc	gga Gly	gee Ala	gct Ala	993
9888 Glu 195	tca Ser	cte Leu	acg Thr	ogg Axg	gaa Glu 200	tgt Cys	aat Asn	att Ile	tgg Trp	agt Ser 205	ati: 1 le	gat Asp	gta Val	ctg Len	aat Asn 210	1041
gca Alia	tta Lou	teg Sor	aaa Lys	gaa Glu 215	CSB Glin	gog Ala	gac ABÇ	cca Pro	aat. Aan 220	ett Leu	gca Ala	tgt Cys	tgt Cys	oca Pro 225	ttt l'he	1089
agu Ser	citti Leu	gal. Asp	ogo Arç 230	tet Ser	Gly 9321	ttt Phe	gta Val	tita Leu 235	gcc Ala	G)II Goa	G)y G(ja	geg Ala	240 Ala 240	gua Val	gtit Val	1137
tgt Cys	ctg Leu	922 Glu 245	aat Asn	tai. Tyr	get: Aep	tua Ser	go: Ala 250	eka Ile	yuy Ala	Arg	GJY GJT	gca Ala 255	acg Thr	att Ile	tta Leu	1185
ace Als	988 Glu 260	att. Ile	aca Lye	ggt Gly	tac Tyr	ясс А.І.а 26 5	caa Gln	tat Tyr	tca Ser	gat Asp	gee Ala 270	gtt Val	aa(; Asn	ua Leu	acc Ther	1233

ogg cos aca Arg Pro Thr 275	gaa gat Glu Assp	att gr Ile Gl 2B0	12. cct lụ Pro	ada Lys	ıle	tta Leu 285	Vje Dol	ata I.le	act Thr	aaa Lys	gee Ala 290	1281
att gag cag Ile Olu Gin	gca cag Ale Gln 295	att to Ile Se	og cog er Pro	Lye .	gat Asp 300	att Ile	yezio Serc	tac T y r	att Ije	aat Asn 305	get Ala	1329
cat 99% ack Mis Gly Thr												1377
asa gca gca lye Ala Ala 325	ctg ggc Leu Gly	Glo 's) Cab ti	et got or Ala 330	tat Tyr	ලාව සෙයි	gte Val	oct Pro	ata Ile 335	tox Ser	agc Sec	ace Thr	1425
ese tot tet Lyb Ser Tyr 340	ace gge That Gly	His L	tt att su Ile 15	gct Ala	ger Ala	gee Ala	99t Gly 350	agt Ser	ttt Phe	gaa Glu	aog Tru	1473
att gta tgt Ile Val Cys 355	gtg ann Val lys	goa ti Ala Lo 360	taget en Ala	gaa Glu	Pen	tge Cys 365	ttg Leu	cca Pro	gca Ala	ace Thr	ttg Leu 370	1521
aat tta cac Aan Leu His	ogg gou Arg Ala 375	gent ox Assp. P.	ca gat ro Asp	tge Cys	gat Asp 380	ctc Leu	aat Asn	tat Tyr	ttg Leu	oct Pro 385	aat Aan	1569
caa cat tgc Gln His Cys	tac acc Tyr Thr 390	get a Ala G	aa cea ln Pro	დად Glu 395	gtg Vel	aca Thr	ctc L≑u	eat Asn	att Ila 400	agc Ser	gca Ala	1617
ggt tte gge	ggg cat	aac go	et geg	ttg	gţt	atc	get	ಶಿಖವು	gta	agg	taa	1665
(i) y Poe (i) y 405	GDy His	Asn A	la Ala 410	Leu	Val	lle	Ala	Іує 415	Vel	Atg		
		ra atg	41 0	at at	it. ga	ua. (%	at to	Lys 415 20 to 20 Se	og aa	at ti	to tet ne Ser	171ê
405	attitig	raatg Met atro	410 gan g Glu A 420 at tat	at at ap Il teg	it. ga e Gli geg	eaaa Langu	kt tg is Tr 42 syn:	Lye 415 20 to Op Se 25	pot a.y. odgan	at ti Xı S i	ne Ser oor	1718 1766
405 ctgatatgtt ç ggg gat ttt Gly Asp Phe	attitig ame eee Amn Pro ema eme	ra atg Met atc c Ile H 435	410 gas gu Glu Au 420 at tat is Tyr	at at mp Il teg Ser	t.gs e Gl geg Ala atg	ua (X IV Hi 298 1978 440 Etg	et ty is Th 42 syn: Ser ser	Lys 415 23 to 25 909 Glu	og aa en Ae tet Sen ete	at ti kn fi ttg Lev tat	ne Ser ogc Ary 445 ota	
ctgatatgtt g ggg gat ttt Gly Asp Phe 430 aat ata cag	abe occ Asn Pro Coa cae Gin Nic 450	met atg Met atr c Ile H 435 eeg gr Pro V	410 gas gr Glu Ar 420 et tat is Tyr tg cag a) Glu	at at ap 11 teg Ser Oga Gly	t. gs e Gl grg Ala atg Met 455	eaa Lys 440 ttg Leu	at to the Th 42 sept: Ser sept Ser	Lys 415 20 to p Se 55 9 mg Glu ttg Leu	og aa er Ar tet Ser ete Leu	at the state of th	oge Ary 445 gta Vøl	1766
Gly Asp Phe 430 aat ata cag Asp The Gln	abc coc Asn Pro coa cac Gin Nis 450 Ut tot Phe Ser 465 gec tot	atr can the Hill the	dio gen ge Glu Au dio et tat is Tyr tg cag a) Glu ta act eu Thr	teg Ser Oga Gly tec Ser 470	t. ga a Gl greg Ala atg Met 455 grt Ala	ease Lycs 4400 ttg Leu ttt	et ty Le Tr 42 segn: Ser segt Ser aca Tho	Type 415 A15 A15 A16 A17 A17 A17 A17 A17 A17 A17	tet ser ctc Leu 992 475	ttg Lev tat Tyr 460 ata Ile	og Ser Ogc Ary 445 Gta Vøl ttg Leb	1766 1814
ggg gat ttt Gly Asp Phe 430 aat ata cag Asn Ile Gln cgg caa cag Arg Gln Gln aac att gat Asn Ile Asp	aac coc Aan Pro caa cac Gln Nie 450 Ut tet Phe Ser 465 gee tet Ala Ser	ra atg Not atr C Ile H 435 ecg q Pro V cas t Gln L ttc c Pha A	dio gas gr Glu A d20 at tat is Tyr tg cag al Glu ta act su Thr gc cag rg Glo 485 se sec	teg Ser Oga Gly tee Ser 470 tat Tyr	t. gs e Gl gcg Ala atg Met 455 gct Als gtt Val	a con the control of	age: Ser age: Ser aca Thr exc	lye 415 33 trops are	og ar er A tot Ser ctc Leu gga Gly 475 tta Leu	ttg Lev tat 9yr 460 ata Ile ccc	og Ser ogr Ary 445 gta Vøl ttg Leb cat His	1766 1814 1862
ctgatatgtt g ggg gat ttt Gly Asp Phe 430 aat ata cag Asn Ile Gln cog caa cag Ang Gln Gln acc att gat Asn Ile Asp 480 caa ctg agg Gln Leu Arg	ame occ Am Pro cma cac Gln dis 450 tot tot Phe Ser 465 goo tot Ala Ser att amt Ile Amn	rs atg Met atr co Ile H 435 com gr Fro V can tr Gln L ttc co Pha A	dio gas gu Glu Au dio at tat is Tyr tg cag al Glu ta ant eu Thr gc cag ng Glu 485 ea Asn ob to ggc	teg Ser GJy too Ser 470 tat Tyr ess	t. gs e Gl gcg Ala atg Met 455 gct Ala gtt Val eccg Tim	a con the case of	et ty age sage sage sage sage age The aca The tye 505	Lye 415 Also trops are series of the series	tct Ser ctc Leu sign 475 ttp. Caa Ciliu aat	ttg Lev tat Tyr 460 ata Ile ccc Pro	og Ser ogc Ary 445 gta Vøl ttg Leu cat His	1766 1814 1862 1910

Glu	ser	Ile	Glu	Авр 530	Trp	Ile	Val	Gln	As p 535	ሕደንገ	C _J /s	Gln	lye	Leu 540	Thr	
ata Ile	aca The	egy egga	920 Glu 545	gaa Glu	gtt Val	tgt Cys	gaa Glu	аад Lys 550	tat Tyr	get Ala	gtc Val	t.t.l: Phe	aga Arg 555	tac Tyr	tat Tyr	2102
ttc Phe	cca Pro	agt Ser 560	gtc Val	act Thr	tet Ser	att Ile	gga Gly 565	tog Trp	tte Phe	ctg Leu	gat Asp	gag Ala S 7 0	ett Leu	gct Ala	ttt Phe	2150
cat Hie	ett Len 575	att Ile	att Ile	aat Aen	tog Ser	aca Thr 580	gga Gly	ttt Phe	ett Leu	aat Aਗ਼	ttt Phe 585	gay Glu	Cac His	tae Tyr	cat Kis	2198
ttt Ehe 590	aac Aen	can Gln	tta Leu	GJu cab	gat Asp 595	tat Tyr	ctg Leu	agt Ser	caa Gln	tet Ser 600	ttt Phe	act Thr	t:tg Leu	cat His	act Thr 605	2246
Gly Ggg	ල <u>ා</u> ා යෲ	gog Ala	att Lle	a a a Lys 610	atc Ilo	Arg Agg	[743 පුතුව	Glu Grug	att Ile 615	gtt Val	æat Aen	agt Ser;	aca Thr	gta Val 620	tta Leu	2294
tta leu	tet 6er	toa <i>Ser</i>	cog Pro 625	gat Asp	ate Ilo	tgt Cys	gtt Val	gaz G19 630	tte Leu	aat Aen	oct Pro	eet Pro	tta Lou 635	ttg Leu	att. Ile	2342
aag Lys	aat Aen	990 Gly 640	gat Asp	aaa Lys	gsac Asp	tat Tyr	att Ile 645	egt Arg	att Ile	lite Phe	tat Tyr	tat Tyr 650	cga Àrg	tgt. Cys	tita Leu	2390
tet Tyr	gat. Asp 655	abe Lys	aga Lys	cet. Pro	att Ile	ttt Phe 660	gta Vəl	tca Ser	ràs 999	act The	tes Ser 665	att lle	atc Ile	tot Ser	aag Lys	2438
atg Met 670	aaa Lye	taa	aag(Jaaa g	ycy a	seetű	уссен	uc ac	ikaag	rtgat	att	tte	ectg			2487
oom	taaa;	gaa t	agoa	atatt	ta at	gato	ووحجز	, ate	tegg	v aga	tgaa	gaaa	ita s	scer _s ;	agagt.	2547
oota	ettt	gt t	toge	ettga	a tt	tgat	agto	e ttg	æcte	ıtet	ggoz	atoo	aa g	ttett.	lytgt	2607
togge	ودوها	yta t	/ggta	att <u>o</u> t	g et	taaa	geeg	aac	ttt	ttc	apat	catt	ct a	ittr:	Adoat	2667
taau	tga;	get o	acto	pacte	t tt	aaaa	tcaa	aat	tgte	utc	tgaz	ittt	ta c	ttee	ittatg	27 27
ttt	titter	toc a	ittes	catt	aaç	aggt	tata	at <u>o</u> X et	aer Aer	gtt Val 675	Leu	gaa Glu	cas Chr	: ggt Gly	680 680	2781
gtt Val	get Als	gct Als	tta Leu	tat Tyr 685	tes Ser	Ala 900	Let: Tyr	tog Ser	690 Gł u	aca Thr	ÇSSS Gl.u	ggt Cly	tet Ser	tog Ser 695	tgg Trp	2829
gtg Val	gga Gly	aac Aan	ttg Ceu 700	tge Cys	tgt, Cys	titt. Sh e	tca Ser	agt Ser 705	gat A s p	yr.a. câtl	gag Glu	cat Kis	tilg Leu 710	cct Pro	att Ile	2677
atc ?le	gtg Val	aat Asn 715	GIY SBB	egt. Arg	egt Arg	tte Phe	ttg Leu 720	att Ile	GJ)) 902	ttt Pie	Val	att Ile 725	cça Pro	gat Asp	cat His	2925
tta Leu	ctt Leu	gust, Asap	eea Lys	acy Thr	gtt Val	ana Ly⁄s	CCC Prv)	ада Агу	gta Vel	tte Pha	gest Asp	ttg : Leu :	gat Nep	atc Flo	aat Asu	2973

- 5 -

730	735	740	
aan caa tit tin cig lys Gin Phe Leu Leu 745	ogt ogt gad dat Arg Arg Asp His 750	ogt gag ata aat att te Arg Glu Ile Asm Ile Ty 755	atett 3021 vrLeo 760
tha ggt gas gga sai leu Gly Glu Gly Asi 769	The Met Asp Arg	con ach aca gat ass en The The The Asp Lys As 770	al ota 3069 an Leu 15
tto gay tta aat gag Phe Glu Lew Ash Glu 780	gat ggt toa dia Asp Gly Ser Leo 785	that att ong acg tta cy Phe Ile Lys Thr Lou Au 790	yo cet 3117 ng Ria
ect ett ogt aan tat Ala len Gly Dys Tyn 795	gtt get att aat Val Ala Ila Asc 800	ort to eact acq cas to Pro Ser Thr Thr (),n P 805	tt atc 31.65 he Ile
tto ttt gca caa gga Phe Phe Ala Gin Gly 810	i aag tac agt gaa • Lye Tyr Ser Gli 815	s tit ato aig ami gor t n Pha lle Met Amn Ala D 820	ta aeg 3213 eu Lys
		tat rga gtc aga att a Tyr Arg Val Arg Ile I 835	
	o ny r ny r Gly Pfe	t gae ott get att ott t 9 Clu leu Amp Ile Leu S 850 8	
aca got taa ttoaca Thr Ala	atat katggagagt (gtt atg gaa 88g 6mm atm Met Glu Lys Lys Ile 860	
Lift acc att gag aa	act gat gac aa		_
She Thr Ile Glu Ly 670	e Thir Asso Asso Ass 87	n Phe Tyr Als Asm Cily A	
670 caa tyt atg yta az	87) a ate tet gtalet	n Phe Tyr Als Asm Cily A	rg His e1. ggt. 3458
670 cam tot ato gte and Gln Cys Met Val Dy 885 gat tog ate eas Ct	87 n atc tot gta et s Ile Aur Val Le 890 m gen ett agt gæ	n Phe Tyr Als Asm Gly A 5 680 t ase cas gos tat agg s u Dys Glm Glm Tyr Arg A	en His en gyt 3458 en Gly et ceg 3506
can tot ato gte and Gin Cys Met Val Ly 865 gat tog ats eas Ct Asp Trp He has Le 900 gtg gog goa tta ag	side tot gta of sile Sor Vails syn m gom oft agt ga u Ala Leu Ser Gl 905 t gat ago ofto ab	n Fhe Tyr Als Asn Cly A 5 680 t ase cas gos tat agg s u Dys Gln Gln Tyr Arg a 895 g get gas sas nga teg s u Als Glu Dys Arg Ser J	en His en ggt 3458 en Gly ett cag 3506 etg cet 3554
can but ate gite and Gln Cys Met Val Ly 865 gat tog ate eas (it Asp Tip He lys Le 900 gite gog gos the acy Val Ale Ale Lem Se 915 tog got tog ace ace	a ate tet gta et s lle Sor Val la 890 a gea ett agt ga u Ala Leu Ser Gl 905 t gat age ete at t Asp Ser Leu Il 920 g aca gat gea ag r 1m Asp Ala Ar	n Fhe Tyr Als Ash Cly A 5 680 t was cas gas tat agg s u Lys Gln Gln Tyr Arg A 835 g get gas acc aga teg s u Ala Glu Lys Arg Ser J 910 a tat gac cas the ass s e Tyr Asp Gln Leu Lys b 925 a act acc tit gat ett g 9 Asn Lys Phe Asp Leu C	en His en ggt 3458 en Gly ett cag 3506 lec Gln etg cet 3554 ett Pro 930 egg tta 3602
can but atu uta an Gln Cys Met Val Ly 865 gat tug ats eas Ct Asp Trp Ile lys Le 900 utu gou uca tta au Val Ala Ala Lem Se 915 tos uut tug aca aca ser Gly Trp Thr Thr 93 tta aat uut uta set tta au val Ala Ala Lem Se 915	a atc tct gta ct s lle &ur Val la 890 a goa ctt agt ga u Ala Leu Ser Gl 905 t gat ag: ctc at r Asp Ser Leu Il 820 g aca gat goa ag r 1mr Asp Ala Ar 5	n Phe Tyr Als Ash Cly A 680 t has con gon tat agg su Dys Gln Glu Tyr Arg A 835 g get gan ann agn teg a Ala Glu Dys Arg Ser J 910 a tat get cas the ann a e Tyr Asp Gln Leu Lys b 925 a ant ann the Asp Leu G 940 t tit All gat gat cag cag cag Phe Tle Asp Glu Gln Y	en His en ggt 3458 en Gly stt cag 3506 le Gln etg cet 3554 et Pro 930 agg tta 3602 ly Leu ets ex:s 3650
can but atu uta an Gin Cys Met Val Dy 865 gat tug ats eas Ct Asp Trp He has Le 900 utu gou uca tta aq Val Ala Ala Len Se 915 too gut tug aca aca sec Cily Trp Tur Tur Tur Tur Tur Asa Giy Val Ty 950 unt cut con usa ga	a atc tct gta ct s lie & Vai la s 900 a gen ett agt ga u Ala Leu Ser Gl 905 t gat ag: etc at r Asp Ser leu Il 920 g acs gat gen ag r 1m Asp Ala Ar 5 l cat get gat ge t tis Ala Asp Al 95 t tis; tige aca ac	n Phe Tyr Als Ash Cly A 680 t has con gon tat agg su Dys Gln Glu Tyr Arg A 835 g get gan ann agn teg a Ala Glu Dys Arg Ser J 910 a tat get cas the ann a e Tyr Asp Gln Leu Lys b 925 a ant ann the Asp Leu G 940 t tit All gat gat cag cag cag Phe Tle Asp Glu Gln Y	en His en gyt. 3458 en Gly ett cag 3506 etg cet 3554 ett Pho 930 egg tta 3602 ely Leu els gta ex: 3650 els gtg cyt 3698

age ara gam two sta sty you saw sty son tit gam get may get gay Ser Thr Glo Tyr Leu Met Alm Lys Met Thr Pho Glu Amp Thr Amp Gly 995 1000 1010	3794
aaa cyc aca tta aca acy aat aty toa ytt gyt gat gaw ytt tit yac lys Ary Thr Jev Thr Thr Ash Mel. Ser Val Cly Asp Glu Val Phe Asp 1025 1020 1025	3842
age aag git tie tie aan gee aft get eet tat gee aft aat aee aat Ser Lye Val Leu Leu Lys Ala Ile Ala Pro Tyr Ala Ile Asm Thr Asm 1030)035 1040	3890
con ttg cat you amu air ant arm ttg ttt gat amm aca goa gog cog Glu leu Nis Glu Aeu Ile Aeu Thr leu Phe Asp Lyo Thr Glu Glu Pro 1045 1050 1055	393B
aca ama too got mot cat cat can ata att amt oft hat ogo tog aca Thr Lys Ser Asp Thr Mis Mis Glin He He Ash Leu Tyr Arg Trp Thr 1060 1065 1070	39B6
the case that cash the age ath chit gam eggs most game agt ach eth cast Leu Pro Tyr His Leu Arg Ille Leu Chu Chy Asn Amp Set Thr Val Amn 1875 1880 1885 1885	4034
aga ata tat gto ott ggt asa gag oca toa eau got aga tto otg aga Arg Ile Tyr Vel Ieu Gly Ies Clu Pro Ser Asn Asp Arg Pho Leu Thr 1895 1180 1185	4082
aga gga agg gta tit aaa oga gga act cat sig iga aigeacgiga Arg Gly Arg Val Phe Lys Ary Gly Thr His Met 1110 1115	4128
taptatatet parameters the beautiful to the second	
taatgigagi ggaggaigig thalggada toottataco gtaactaiic oggacaogca	4188
nergadast assatante statescada aratastace desertata ealestatan	
	4248
pettyrinet gaagigriic aigigacagy gigtirytyy acgugigyti akiakyatyy	4248 4308
pothyripot gaagigriic aigigacagy gigiirgigy acquatgyti alitalgatgy atalraigat gicacaatca tigalaacia cygiiglosg cataaattia gaallictic	4248 4308 4368
gottgetisct gaagtgette atgtgacagg gigtingtog conggiggit attatgatgg ataleatgat gicacaatea tigataaria nygitigloog cataaattia gaattiette ggittaatatt ggaogtgoge taagrafiage gagdataagi tgattiteek kagtanaaa	4248 4308 4368 4428
politicio gaagigette aigigacagi gigitingigi acquaggit altaigatgi alaicaigai gicacaatca tigalaacia cygitigicag cataaatti gaatiteette ggitaatatt ggaogigoge laagcalage gagaataagi tgalliteek lagtamaaa cettigitta tgebggiaaa cycatgigog tikgooagoa attaatatat (contacto	4248 4308 4368 4428
portogetost gaagtgette atgigacagg gigitingtog segugiggit attatgatog ataleatgat gicacaatca tigataacia cygrigicog cataaatti gaattiette ggitaatati ggacgigoge taagcaloge gagaataagi igattiiteel lagianaaaa cettigitta igeograaa cycaigigog tikgcoagca attaatata loosattatig aastaggaat atageratat eigiaattat acataaacga attittaete gaataacti tizatigate aaseggaaa titaaa aig aaa gel ace gat ata tat ico aat Met Lys Ala The Asp Ile Tyr Ser Aan	4248 4308 4368 4428 4488
gottgetget gaagtgette atgtgacagg gigtingtgg cogngiggtt attatgatgg atateatgat gicacaatca tigataacia nygitigtoog cataaattia gaattiette ggitaatati ggaegtgoge taagrashge gagantaagi tgattiteek tagtamaaa cettigitta tgetggiaaa ogcatgigeg tikgoongon attaatatai toomitatig aaataggaat atageestat eigitaattat acataanega attittaete gaatataatt tiantigate amacaggaaa titaaa aig aaa gek see gat ata tat too aat Met Lys Ala Tim Asp Lie Tym Ser Asm 1120 1125 get tit aat tie ggi tet tat att aat aek ggi gie gat eer aga aca Ala Pha Asp Phe Gly Sar Tym Lie Asp Tim Gly Val Asp Pro Ang Tim	4248 4308 4368 4368 4428 4988 4541
gottgetget gaagtgette atgtgacagg gigtingtgg cogngiggti attatgatgg atateatgat gicacaatea tigataania nggitgbog cataaatita gaatteette ggitaatati ggacgigoge taagrafage gagaateagi igattiteen tagtaabaaa cettigitta igetggiaaa ogcaigigog tikgocagon attaatatat toomitaitg aaataggaat atageestat cigitaatiat acataaacga attittacie gaatabaati tiantigate aaacaggaaa titaaa aig aaa gel ace gat ata tat ice aat Met Lys Ala Tim Asp Ile Tyr Ser Asn 1120 1125 get tit aat ite ggi tet tat ait aat aci ggi gie gat oce aga aca Ala Pha Ash Pha Gly Sar Tyr Ile Ash Tim Gly Val Asp Pro Arg Tim 1130 1135 2140 ggi caa tat agi gea asi ali, asi ali ate acg ita aga cel aat aat Oly Gln Tyr Ser Ala Ash Ile Ash Ile Ile Tim Leu Arg Pro Ash Ash	4248 4308 4368 4428 4428 4541 4541

		Ile			crt Leu		Phe					Gly				4781
	Сув				ಕ್ರಮ ಆಯೆ	Pro					Le∪					4829
lys	aaa Lys 225	cta Leo	aaa [ys	gat Asp	ttg Leu 1	oge Arg 1230	gta Vel	tat Tyr	aag Lye	Leu	gst ኦ ቋን 1 23 5	Ser. ago	aat Aan	act Thr	ttt Hos	4677
tat Tyr 1240	W.1	tat Tyr	arc Arn	Lye	aac Aen 1245	GJÀ ð3c	att Ile	ala Ile	Glu	ata Ile 1250	ett Leu	оад Гув	oga Arg	Ile	990 GLy 1255	4925
teg Ser	agt. Sor	gat Asp	Iļe	ពូចរ សំខ 1260	а а а Цув:	aca. Thr	gtt Val	Ala	ctt Leu 1265	Gju Gera	ttt Fhe	cet Pro	Āф	ggt Gly L270	gsa Glu	4973
		AΞp			tat Tyr		Ser					Ser				S 02 1
tac Tyr	Αιτιχ	gtg Val L290	ace Thr	ggt Gly	aga Lys	Thr	tet Tyr 1295	ett Læu	aaa Lys	ete Leu	Asm	tac Tyr 1300	t ct Ser	gga Gly	aat Asn	5069
Am	tgt Cys 1305	aca Thr	tca Ser	yal Yal	gee Glu	tec Nyr 1310	cet kro	gat Asp	get Asp	Aso	ast Aen 1315	att Ile	tet Ser	Υ)a Θοά	aan Dye	5117
ata 11e 1320	ΑÌü	ttc Phe	gat Nep	Тух	egt Mrg 1325	аас Дул	gat. Asşo	tac Tyr	144.1	att 11e 1330	acy The	gtg Val	act Thr	Val	cot Pro 1335	5165
tad Tyr	gat Asp	get Ala	ser	99t Gly 1340	oct Pro	att Ile	gat Asp	Ser	900 Ala 1345	cga Arg	ttt Phe	aæg Lye	Met:	acc 1 55 1350	tat Tyx	521 3
cəg Gln	ace Thr	Leu	222 Lys 1355	Gly	gta Val	¢ti: Phe	БLО	gtt Val 1360	atr Ne	agc Ser	æ.c Thr	Fhe	ogc Arg 1365	aca Thr	oca Pro	5261
acc Thr	Gly	tat Tyr 1370	gtt Val	<i>ලාග</i> සෞඛ්	ctg Leu	Val	agt Ser 1375	tat Tyr	aaa Lys	ලෝ ෆ බෞයි	Aso	999 Gly 13 8 0	cat His	aaa Lys	Vəl Vəl	5309
Tim	gac Asp 1385	acg Thr	gæa Glu	tat Tyr	att Ile	cet Pro 1390	Tyr	gog Ala	get Ala	Ala	ete Leu 1395	act Thr	at.t Ile	CHA Gli	coc Pro	5351
992 Gly 1400) Deal	ejy. Ogs	cza Gln	iro	gng Ala 1405	gtic Val	egc Ser	age Lye	Ser	tat Tyr 1410	gaer Glu	tat Tyr	agt Ser	Seri	gta Vai 1415	5409
cat His	aac Asn	tte Pho	Leu	gga Gly 1420	tat Tyr	tet Ser	tict Ser	Gly	099 Arŋ 1 42 5	Thr	agc Ser	ttt Phe	аер	toe Ser 1430	Sér Sér	5453
caa Gln	gat Asp	ASTI	ttg L c u 1435	tat Tyz	ttg læi	gtc Val	Terr	999 Gly 1440	Lys	tac Yyr	act. Thr	IJYY	եմա Ser 14 4 5	Ser	att Ile	5503
gas	ogg	gţţ	tte	get	ggt	CBB	agt	gtg	gtt	ton	gta	ata	gan	cga	gta	5545

GIu	Arg	VBI 145D	Leu	Asp	GIĀ	Glu	Ser 1455		Val	Ser		Ile 1 4 60		Aाग्न	Wal	
Pha	aat As n 1 46 5	Į.,	ttc Phe	cat Nie	Leu	atg Met 1 47 0	Thr	lys aaa	Gju Gan	Ala	886 Lye 1475	aca Thr	caa Glin	gat Asp	aat Aen	5597
ಕಿತವು 199 1480	Arg	att Ile	ace Thir	Thu	928 Glu 1485	att Ile	act Thr	tec Tyr	Aan	gag Glu 1490	gat. Asp	cra Leu	tca Ser	Liys	agt Ser 14 9 5	5645
tto Phe	bça Ser	gag Glu	$a_{\rm L}$	сса 2то 1500	Glo	aat Asn	tta tta	Gin	caa Gla 1505	ect Pro	tet Ser	ege Arg	Væl	tta Leu 1510	exx: Thir	5693
æt Arq	tat Tyr	The	gat <i>As</i> p 1515	ata Ile	ടേക ദിഥ	aca Thur	Ασι	act Thr 1520	tca Ser	oga Arg	gaa Glu	Glu	act Thr 1525	gt¢ Val	ast: Aem	5743
att Iìc	Ly	agt Ser 1530	yeb	gat Asp	tgg Trp	gga Gly	aa t Aen 1535	act Thr	eta Leu	ctt L⇔,	Ilc	act Thr 1540	gag Glu	acc Thr	agt Ser	5789
CIÀ	ata Ile 15 45	ලුවා සෘති	aae Lys	gae Glu	Tyr	gtf. Val 1550	tat. Iyr	tat Tyr	eeg Pro	Val	aat Aan 1555	ejà ac	gaa Glu	ej åår	aat Jen	5837
agt Ser 1561	CA≅	ect Pro	gcc Ala	yab	ece Pro 1565	ttg Leu	ggt Gly	ttt Yhe	Ser	09 9 Arg 1 5 70	tte Phe	tta Leu	aza Lys	Ser	gtt Val 1575	5885
aog Tik	cee Gln	FA2 888	GIA	teg Sør 1580	ect Pro	get Asp	get Ala	Ala	caa Gln 1585	agt. Ser	etc Val	gen Ala	Ae ty	ваа Lys 1590	Vel Øt©	5933
att Iie	cat His	Tyr	aca Thr 1595	tat Tyr	Gln Gln	aaa Lye	Fhe	eet Pro 1600	act Thr	ttt She	acc The	Gly	get Ala 1605	tat Tyr	gti: Val	5981
਼ਿ ਹ	GIII	tat Tyr 1610	gtc Val	agt Ser	aaa lys	gtc Val	tca Ser 1615	Gjn ðsíð	acg Ther	ata Jl⊕	Asp	aat Asn 1620	ees Lys	ata ile	gog Ala	6029
Arg	ace Thr 1625	ttt Phe	ser ser	tat Tyr	ns1	aac Asn 1630	toa Ser	ccg Pro	acg Thr	Ser	aaa Lys 1635	tet Sar	cet His	ggt Gly	teg Ser	6073
tta Leu 1.6 4 0	WIN.	rae Pae	eta Ile	Thr	tca Ser 1645	gtg Val	atg Met	eat: Azn	ABID	cag Gln L650	can Gln	acg Thr	gte Vel	Thr	aca Thin 1655	61.25
ttt Bhe	ees Lys	tat Tyr	GIU	tat 'Jyr 1660	tçe Ser	gaa Glu	egt. Ser	Gla	atg Met 1665	acc Thr	aca Thor	aat Asn	Ala	acg Thr 1670	gtg Val	6173
ace Thr	ggt Gly	Phe	gat Aep 1675	Gly ශූල	ala Ala	cat His	X+1	ç saa Glu L680	teg Sor	aaa Lys	aat. Asto	Val	acy Thr 685	tet Ser	att Ile	6221
tat Tyr	TIT	cat His 1690	c g g Arg	Gln cae	ett Leu	ogt Arg)	ааа Lys 1695	gtt Val	gat Asp	gta Vel	A/AII	cac Ris 1700	gtg Val	att. Ile	acc Thr	6269
gat Asp	cag Gln	tet Ser	tet Tyr	gst Asp	ctt Leu	ttg Leu	ggt Gly	ogo Ang	att Ile	aca Thr	999 Glv	cea Glr	att lle	att Ilo	gat Aso	6317

1705	1710	1715		
ooc ggc acg gca Pro Gly Thr Ala 1720	ege gee att as Arg Glu lle ly 1725	a ogt aat tac gtt s Arg Aso Tyr Val 1730	tat cas tak con Tyr Ghn Tyr Pro 1735	6365
Oly Gly Asp Glu	aat gel CUL tg Aso Asp Phe Tr 1740	g ccg gtg afg ata p Pro Val Met Ile 1745	gam gtt gat tot Glu Val Asp Ser 1750	6413
caa gge gte aga Gln Gly Val Arg 1.755	cyt eas and na Ang Lys Thr Hi	nt tec get ggs elg a Tyr Asp Gly Met 1760	gge ogt att tgt Gly Ang Ile Cye 1765	6461
tog act gee gee Ser Ile Glu Glu 1770	cea gat gat ga Gln Asp Asp As 177	nt gar ger tog age op Oly Ala Tro Gly '5	acs toy ggg att 'Mmr Ser Gly Ile 1760	6509
tat cas ggr aca Tyr Gln Gly Thr 1785	tat oga aaa gt Tyr Arg Lys Va 1790	t ett gee aga cam il lew Al¤ Arg Gin 1795	tat gat git tig Tyr Amp Val Leu	6557
ggg cwg ttg agc Gly Gln læu Ser 1800	aag gaa att to Lys Glu Ile Sa 1805	a aat got tog tta or Aen Aep Tro Lev 1810	tgg aat the tet Trp Asn Leu Ser 1815	6605
Ala Aso Pro Leu	gtt ogt ett ge Val Arg Leu Al 1820	el axy cog tbg gtt a Thr Pro Leu Val 1825	ace acy asa acc Thr Thr Lys Thr 1830	6653
tat ana tat gat Tyr Ly s Tyr As p 1635	Gly Trp Gly As	at out lan ago acg on Leu Tyr Ser Thr 1940	gaa tac agt gat Glu Tyr Ser Asp 1845	6701
ggt ogg ata gag Gly Arg Ile Glu 1850	ctg gas atc ca Leu Glu Ile Pi 185	it gat ect att aeg <i>a ha</i> p Pro Ile Tur 5	agg see att act Arg Thr Ile Thr 1860	6745
cas 99g gizt aas Gln Gly Val bys),865	gge the ggg at Gly Leu Gly Ma 1870	g lia asl att oxg et leu Asm Ile Gin 1875	can aat aat ttt Glo asa Asa Phe	679.
gang cama cog get Glu Glu Pro Ala 1880	tog ato aaa go Ser lie lys Al 1885	et gtg tat oet gat la Val Tyr Pro Asp 1890	ggt acg ata tal Gly Thr lle Tyr 1895	5845
Ser Thr Arg Thr	tat ogt til ga Tyr Arg Tyr Ae 1900	at gges ttt ggst ogst ap Gily Phe Gily Arg 1905	aca gtg acg gas The Val The Glu 1910	6 8 93
aca gat gca gea Thr Asp Ala Glu 1915	Gly His Als Th	x caa all gga tot x Gln Ile Gly Tyr 1320	gat gtg tit gst Asp Val The Asp 1925	6941
ogt ata gtg aaa Arg Ile Val Ivs 1930	aaa acg ttg co Lys Thr Leu Pr 193	ca gar gga aca ata no Asp Gly Thr Ile 35	tta gea too got Leu Glu Ser Ala 1940	6989
tat gca age tit Tyr Ala Ser Phe 1945	age cat gaa ga Ser His Glu Gl 1950	ea tta att tog goe Nu Lou Iho Sor Aln 1955	Len Asn Val Asm	7033
ggo ace cag ttg Gly Thr Gln Leo 1960	gyg gos tta gt Oly Ala Lou Va 1965	it tat gat ggt ett. 네 ፕ૪૯ አድር Gly Leu 1970	999 099 sta ata Gly Arg Val Ile 1975	7089

agt gal Ser Asp	. ecg Thr	Val	1980 Gly ggt	g gt Gly	oge Arg	aaa Lys	The	ന്ദ്രമ ദിധ 1985	tat Tyr	tta Leu	tat Tyr	Gly	eet Pro 1990	cza Gln	7133
ggt geo Gly Asy) lys	cxxg 1995	11e	cag Gln	tca Ser	Υlσ	act Thr 2000	eet Pro	ting Ser	cat Ris	A21)	രമു Lys 2005	caa Gìn	eat Aen	7381,
atg gat Met Asg	tac Tyr 2 0 10	ete Leu	tac Tyr	tat Tyr	Len	ggt aly 2015	aqt Ser	gtg Val	atg Met	Ser	eaa Lys 2020	ttt Rhe	occ The	acg Thr	7229
999 not Gly Tha 2029	: Авр	ræe Gln	cea Gln	ÀETI	titit Phe 2030	egt Ang	tat Tyr	cat His	Ser	аа а Ly s 2035	acg Trư	Gly Gly	aca Tu	tta Leu	7277
tta tet Leu Seo 2040	gcg Ala	toe Ser	alg	990 Gly 2 04 5	gta Val	let Ser	Gju cañ	Thr	aat Aan 2050	tae Tyr	agt Sor	tat Tyr	Phe	cca Pro 2055	7325
tog ggl Ser Gly	: 91& 'Val	Ten	تھی Gln 2060	args Arg	gea Glu	(ucas Sear	Phe	tta Leu 2065	cgg Arg	gat Asp	aat Asn	Taro	cog Pxo 2 07 0	att Tle	7373
tos tog Ser Sei	: Cly	gag Clu 2075	tac Tyr	ett Leu	tat Tyr	Thr	atg Mat 2080	toc Ser	ggt Gly	tig Leu	Ile	caa Gln 2085	ogt Arg	cat Nis	7421
aas gat Lys Asg	. agt . Ser .2090	ttt Pix	ggt Oly	c at Hís	Asn	cet His 2095	gtt Val	tet Tyr	agt Ser	Tyr	gant Asp 2100	gct Ala	cag Gin	egy Özen	7469
aga ttg Arg Len 2109	ופע נ	эма Lyre	aca Thr	Glu	саў GJn 2110	gat Asp	gca Alle	cae Cln	Tyr	get Ala 2115	aca Thr	ttt Phe	Glu	tat Tyr	7517
gac aat Asp Ast 2120	gtt Val	ggg Gly	yrd	t.tg Leo 2125	ata Ile	aco Thr	acg Tha	Thr	acc The 2130	aaa lys	gac Asp	acy Thr	Tim	tca Ser 2135	7565
tta ber Leu Ser	ces Cln	Leu	9kg Val 2140	aca Thr	aaa liye	atc Ile	Glu	tat Tyr 3145	gat Acy	get Ala	ttt Libe	Asp	093 Arg 2150	gsa Glu	7613
ata aaa Ile Lyx	arg	tog Ser 2155	cta Lui	ett Ile	agt Ser	ÆP	tte Phe 2160	tca Ser	ata Ile	caa Gln	Val	att Iie 2165	acc Thr	tta Iæj	7661
age tat Ser Tyr	acg Thr 2170	aag Lys	aat A sn	eat Asn	Gln	atc Ile 1175	egt Ser	GJN CBB	egl; Arg	Ile	acc Tim 180	tee Ser	ate Ile	OST OST	7709
999 gtg Gly Val 2169	. עבו	alig Met	eaa Lys	yezi	gaa Glu 2190	ogt. Arg	tat Tyr	caa Oln	Tyr.	gat Asp 2195	aat Asn	aat. Aan	caa Gìn	ogo Arg	7757
tha ago Leu Sei 2200	caa Gln	tæ; Tyr	Gln	tgt Cys 2205	G)ព ខ្មែរពិ	ssia Gly	Gjů Gas	GJrs	tet Sur 210	eeg Pro	att Ile	gat Asp	His	acg Thr 215	7805
ggt ogt Gly Arg	gta Val	Гап	ваt Asn 2220	GJU CHB	eju cas	att Ile	Туг	cat His 2225	tat Tyr	7 20 7 20	caa Gln	TY	67A 67A 5330	aat Asm	7853

all any cyg ctr gat ant ack this cya gat gyt may gam acy gig ga The Lya Ary Low Asp Ann Thr Tyr Ary Asp Cly Lys Clu Thr Val As 2235 2240 2245	
tat cat the agt cas god gat cos act cas cft att ogt att see ag Tyr Nis Phe Ser Clin Als Asp Pro Tir Glin Leu lle Arg Ile Tir Se 2255 2260	e 7949 r
ger aan dag dag ata gag tia agi tiat get get aat gge aac dia ac Aap Lys Olm Gin lie Giu Leu Ser Tyr Asp Ala Asm Gly Asm Leu Th 2265 2270 2275	a 7997 r
ogt gac gam amm ygg cam acg oto att tac gat cam mat mat cgo tt Arg Asp Glu Lys Gly Gln Thr Lou Ile Tyr Asp Gln Asn Asn Arg Le 2280 2285 2290	ũ
gta can gto and get ong tig ggo aat cin gig igo ago iac ceg ta Val. Gla Val Lya Asp Arg Leu Gly Asa Leu Val Cys Ser Tyr Gla Ty 2300 2305 2310	t. 11093 T
got god the eac saa tha acc gra can git the gog ast get see gr Asp Ala Leu Asm Lys Lou The Ala Gim Val Leu Ala Asm Cly The Va 2315 2320 2325	a
aat oga oag oot tat got tot go! aas gtg ang aat att oos ttg gg Asn Ang Gln His Tyr Ala Ser Gly Lys Val Thr Asn Ile Gln Leu Gl 2330 2335 2340	У
gat gas gag att act tog ttg age agt gat aag caa oga att gga ca Asp Glu Ala Ile Thr Trp Leu Ser Ser Asp Lys Gln Arg Ile Gly Hi 2345 2350 2355	E
can agt gar ang ant ggt cma tam gto the that cma that ggt att go Gin Ser Ala Lys Asm Oly Cln Ser Vai Tyr Tyr Gir Tyr Giy 11e An 2360 2365 2370	P '5
cat aac agt acg git atc god agt cag amc gam amc gag itg aig go His Asn Ser Thr Val Ile Ala Ser Gln Amn Glu Amn Glu Jew Met Al 2380 2385 2390	a
tta too tat aca cot tat ggo tit agg agt tita att too toa tita oo Leu Ser Tyr Thr Pro Tyr Gly Phe Arg Ser Leu Ile Ser Ser Leu Pr 2395 2400 2405	0
ggt tig aat gge gea eag git got ees git aca gge tog tae tie ti Gly Leu Asn Gly Ala Gln Val Asp Pro Val Thr Gly Trp Tyr Pha le 2410 2420	D.
synt oan soa tak ook okk the ase cong oft etc atg agg tir can ag Gly Amn Gly Tyr Arg Val Phe Amn Pro Val Leu Met Arg Phe Ris sa 2425 2430 2435	rc 8477 r
cor gat agt tgg agt oot tit ggt ogg gga ggg att aac cet tat ac Pro Asp Ser Tno Ser Pro Phe Gly Arg Gly Gly Ile Asm aro Tyr Th 2440 2465 2450 2450	α
tat type caa gype gat eee ata aac eyy; att gat etg aac yy); cat et Tyr Cys Gln Gly Asp Pro Ilo Asp Asp Ilo Asp Leu Asn Gly His Le 2460 2465 2470	£ 8573 u
agt god ggd ggg ata tha ggd att grg ota ggg gda att ggd atm at Ser Ale Gly Gly Ile Leu Gly Ile Val Leu Gly Ale Ile Gly Ile Il 2485 2480 2485	t 8621 o
yte ggg att gta tea etg ggm geo gga geg geg att age geg ggt et	c 86 6 9

val Gly 11e val Ser Leu Gly Ala Gly Ala Ala 11e Ser Ale Gly Leu 2490 2495 2500	
att grit geg gag gar get tig gag gog att get tet ace age geg ett i The Ala Ala Ghy Ghy Ala Leu Ghy Ala Ile Ala Ser Thr Sar Ala Leu 2505 2510 2515	B717
gra git act gog act gic att ggn tig get gee gni tog ata ggg att 4 Ala Vai Thr Ala Thr Val lie Gly Leu Ala Ala Asp Ser lie Gly lie 2520 2535	8 76 5
grow ton gon gon the ing gas ase get ong ase are tot ggg ata the Als Ser Als Als Leu Ser Glu bys Asp Pro Lys Thr Ser Gly Ile Leu 2540 2550	8813
ast typ att agt gog gog tog gog gtt tta agc ttt ggt ate age gon 8 Asm Typ Die Ser Ala Gly Leu Gly Val Leu Ser Phe Gly Die Ser Ala 2555 2560 2565	8861
ata acc tit acc int ing ctg gia ass tog gin ogs agt ggit tot cag The Thr She Thr Ser Ser Leu Vol Lys Ser Ala Ang Ser Gly Ser Gln 2570 2575 2580	B909
gra gtc age geg ggt gtt ate ggg tea gtg cet ett gas til ggt gas (Ala Val See Ala Gly Val Ile Gly Ser Val Pro Leu Glu Phe Gly Glu 2585 2590 2595	8957
gtt get age ogt tee age age ogs tgg gat att geg tha tet teg aba 9 Val Ala Sar Arg Ser Sar Arg Arg Tro Asp Ila Ala Leu Ser Ser Ile 2600 2605 2610 2615	9005
tog tig gge gea aat geg geg tot oto tot aog ggg ata geg geg geg Sar Leu Gly Ala Aso Ala Ala Ser Leu Ser Thr Gly Ile Ala Ala Ala 2620 2625 2630	9053
gog git gea gac agi aai geg aai gea gel aai ai; eig gga igg gia g Ala Val Ala Asp Ser Aso Ala Aso Ala Aso Ile Leu Gly Trp Val 2635 2640 2645	9101
too tit ogt tit ogt ges gis tog ace and the ggs ate also gay oft. Ser The Gly Phe Gly Ale Val Ser The The Ser Cly 12e Ile Glu Leu 2650 2655 2660	9149
acg tgt acg get tat ges gtg sat cet cag act tgg gas ctg agt toa 9 Thr Arg Thr Ala Tyr Ala Val Ash Bis Gln Thr Trp Glu Leu Ser Ser 2675 2670 2675	3197
toa goa ggt act bog gag gan gtg aag oot ata egt tgt ete gti tea g Ser Ala Gly Thr Ser Glu Glu Val Lys Pro Ile Arg Cys Leu Val Ser 2680 2685 2690 2695	9245
cae oge tog aat ong ang ong tyn atyttanoer terteggyen gittgegtten 9 His Ary Tro Aso Glo Nys Clo 2700) 299
tomaacgitt egamatagia eegygmaets titageemat egicembiga anoocgiami s	
gtyttgegac gtogtttgac aatalaaags ttetgegaan egaktegtta agtetemoga s	
aaaataacta ttaggegaca tttgogtogc cttttttaag geactttate aggttacatt 9	
falsagaage tattitigitt tegseggstg tiggittete tegsestaman matagaggs 9	
asignigica aggytysias typitasity tamasiaty; patattatic positiatai 9	599

greentgias tiectritai taittaatti taitgeatti getaegegaa alegeriiai 9659 aattiratti tiaatasatt attattesi esitaaseta aaatasatta tiictags 9717

<210> 2

<211> 417

<212> PRT

<213> Photorhabdus luminescens

 $\sim 4000 \sim -2$

Met Gin Ang Ala Gin Ang Val Val Ile Thr Gly Met Gly Ala Val Thr 1 5 10

Pro Ile Cly Clu Asp Val Clu Ser Cys Trp Gln Ser Ile Ile Glu Lys 20 25

Gim His Arg Fine Ris Arg Ile Glu Phe Pro Asp Ser Phe Ile Asm Ser 35 40 45

Arry Phe Pive Ser Phe Leu Ala Pro Asn Pro Ser Arg Tyr Gln Leu Leu 50 55 60

Pro Lys Lys Leu Thr His Thr Leu Ser Asp Cys Gly Lys Ala Ala Leu 65 70 75 80

bys Ala Thr Tyr Gln Ala Phe Thr Gln Ala Phe Gly Val Asn Ile Sar 85 90 95

Pro Val Glu Tyr Tyr Asp Sys Tyr Glu Cys Gly Val Ile Leu Gly Ser 100 105 110

Gly Top Cly Ala Ile Asp Asm Ala Cly Asp His Ala Cys Gln Tyr Lys 115 120 125

Gin Ala Lys Leu Ala His Pro Met Ser Asa Leu Ile Thr Met Pro Ser 130 140

Ser Met Thr Ala Ala Cye Ser Ile Met Tyr Clly Leu Arg Gly Tyr Gln 145 150 155 160

Asm Thr Val Met Ala Ala Cys Ala Thr Gly Thr Met Ala Ile Gly Asp 165 170 175

Ala Phe Ghu Ile Ile Arg Ser Gly Arg Ala bys Cys Met Ile Ala Gly 180 185 190

Ala Ala Glu Ser Leu Thr Arg Glu Cys, Asm Illo Trp Ser Ille Asp Val 195 200 205

Iou Asn Ala Leu Ser las Glu Gln Ala Asp Pro Asn teu Ala Cys Cys 210 220

Pro Pho Sur Leu Asp Ang Ser Gly Pho Val Leu Ala Glu Gly Ala Ala 225 235 240

Val Val Cys Leu Glu Asn Tyr Asp Ser Ala He Ala Arg Gly Ala Thr 245 250 255

The Leu Ala Glu Ile Lys Gly Tyr Ala Glu Tyr Ser Asp Ala Val Asp 260 265 270

Less thr Arg Pro Thr Glu Amp Ile Glu Pro Lym Ile Ieu Ala Ile Thr 275 280 285 lys Ala The Ghu Ghu Ala Ghu Ile Ser Pro Lys Asp Ilo Asp Tyr ile 290 295 300

Aan Ala Mis Gly Thr Ser Thr Pro Lou Asa Asp Leu Tyr Glu Thr Oln 305 310 320

Ala Ile lys Ala Ala Lau Gly Gln Tyr Ala Tyr Gln Val Pro Ile Ser 325 330

Ser Thr lys Ser Tyr Thr Gly His Lew Ile Ala Ala Ala Gly Sor Phe 340 350

Glu Thr Ile Val Cys Val Lys Ala Leu Ala Glu Asn Cys Leu Pro Ala 355 360 365

Thr Leu Asn Leu His Arg Ale Asp Pro Asp Cys Asp Leu Asn Tyr Leu 370 380

Pro Aen Glu Idis Cys Tyr Thr Ala Glu Pro Glu Val Thr Leu Aen Ile 385 390 395 400

Ser Ala Cly Phe Gly Gly Ris Asa Ala Ala Lou Val Ile Ala Lys Val 405 410 415

ALT]

<210> 3

<211> 253

<212> PKT

<213> Photorhabdus lundnescens

<**40**0> 3

Met Glu Asp Ile Glu His Trp Ser Asn Phe Sor Gly Amp Phe Asn Pro 1 5 10 15

The Ris Tyr Ser Ala Lys Ser Glu Ser Too Arg Aso The Glu Gin His 20 25 30

iro Val Gin Gly Met Leu Ser Leu Leu Tyr Val Arg Gin Gin Phe Ser 35 40

Gin Lev The See Ala Phe The The Gly Ile Lev Ash Ile Asp Ala Ser 50 55

Phe Arg Gln Tyr Val Tyr Thr Ala Leo Pro Kis Gln leo Arg Ile Aso 65 70 75 80

The Lys Aso Lys The Phe Lys Leo Glo Aso Pro Sar Lys Glo Aso The 95 99

Leo Phe Cly Asm Thr Ser Val Glu Asm Thr Met Glu Ser Ile Glu Asp 100 105 110

Trp Ile Val Gln Asp Asn Cys Gln Lys Leu Thr Ile Thr Cly Glu Glu 115 120 125

Vei Cys Glu Lys Tyr Ala Val Pho Arg Tyr Tyr Phe Pro Ses Vei Thm 130 135 140

Ser Ile Cly Tro Phe Leu Asp Ala Leu Ala Phe His Leu Ile Ile Asm 145 150 155 160

Ser Thr Gly Phe Leu Asm Phe Glu His Tyr His Phe Asm Glm Leu Glm 185 170 175 And Tyr Leu Ser Glin Ser Phe Thr Leu Kis Thr Gly Glin Ala Ile Lys 180 1.85 190

Ile Arq Lys Glu Ile Val Asn Scr Thr Val. Leu Leu Ser Ser Pro Asp 195 200 205

The Cys Wal Glu Leu Asn Pro Pro Leu Leu Ile Lye Asn Gly Asp Lys 210 220

Amp Tyr Ile Arg lie Whe Tyr Tyr Arg Cym Leu Tyr Amy Lym Pro 225 230 230

lle Mhe Val Ser Lys; The Ser Ile Ile Ser Lys; Met; Lys. 245 250

<210> 4

<211> 186

<212> PRT

<213> Photorhabdus luminescens

<400>-4

Wet Asn Val Leo Glo Glo Gly Lys Val Ala Ala Leo Tyr Ser Ala Tyr 1 5 10 15

Sext Olio Thr Glio Gly Serr Serr Trip Val Gly Ash Lew Cys Cys Phe Ser 20 25 30

Ser Asp Arm Glu His Leu Pro Ile Tle Val Asm Gly Arg Arg Phe Leu 35 40 45

The Ghu Phe Val The Pro Asp His Leu Leu Asp Lys The Val Lys Pro 50 55

Arg Val Phe Asp Leu Asp Ile Asp Lys Glo Pho Lou Leu Arg Arg Asp 65 70 75 80

His Arg Clu Ile Asn Ile Tyr Leu Leu Cly Clu Gly Asn Phe Met Asp 85 90 95

Arg Thr Thr Thr Asp Lym Asn Leu Phe Glu Leu Amn Glu Asp Gly Ser 100 105 110

Leu Phe Ile Lys Thr Leu Ary His Ala Leu Gly Lys Tyr Val Als Ile 115 120 125

Asn Pro Ser Thr Thr Glo the Tle the the Ala Glo Gly byz Tyr Ser 130 135

Glu Phe Ile Met Asn Ala Leu Lys Thr Val Glu Asp Glu Lou Ser Lys 145 150 155 160

Nrg Tyr Arg Val Arg Ile Ile Pro Glu Leu Gln Gly Pro Tyr Tyr Gly 165 170 175

Phe Glu Leu Asp Ile Leu Ser Ile Thr Ale 180 185

<210> 5

<211> 258

<212> PRT

<213> Photorhabdue luminescens

<400> 5 Met Olu Lys Lys Ile Thr Thr Phe Thr Ile Glo Lys Thm Asp Asp Asn Phe Tyr Ala Amn Gly Arn His Gln Cys Met Val Lys Ile Ser Val Leu Lya Glu Glu Tyr Arg Axu Gly Axu Trp Ho Lya Leu Ala Leu 35 40 Ser Glu Ala Glu Lys Arg Ser Ile Gln Vel Ala Ala Leu ser Amp Ser 50 60 Leo lie Tyr Asp Gln Leo Lys Met Pro Ser Gly Trp Thr Thr Thr Asp Ala Ang Asn Lys Phe Asp Len Gly Len Len Asn Gly Wal Tyr His Ala Asp Ala Phe Ile Asp Glu Gln Val Thr Asp Arg Ala Gly Asp Cyr. Cyr. 95 100 105 130 Thr Asn Ghu Aen Tyr Gln Aen Ser Val Lys Ser Val Pro Glu Yle Yle Tyr Arg Tyr Val Ser Ser Asn Arg Thr Ser Thr Glu Tyr Leu Met Ala Lys Met Thir Pixt Glu Asp Thir Asp Gly Lys Arg Thir Leu Thir Thir Ash 145 150 155 Met Ser Val Gly Asp Ghu Val Phe Asp Ser Lye Val Leu Leu Lys Ala lle Ale Pro Tyr Ala Ile Asn Thr Asn Gln Leu Wis Clu Asn Ile Asn 185 The Lew Pic Asp Lys The Clu Olo Pin The Lys Sor Asp The His His 195 200 205 200 Glo lle lle àsa Leu lya Arg Tro Tar Leu Pro Tya His Leu Arg Ile Low Olu Gly Asn Asp Ser Thr Vel Asn Arg Ile Tyt Vel Leu Gly Lyz 225 230 235 Glu Pro Ser Asn Asp Arg Phe Leu Thr Arg Gly Arg Vel Phe Lys Arg **24**0 Gly Thr His Met 255 <27,0> 6 <211> 1584 <212> PRT <213> Photorhabdue luminescens

- Amm lie lie Thr Leu Arm Pro Amm Asm Val Gly Asm Ser Glu Cln Thr 35 40 45
- Leu Ser Leu Ser Phe Ser Pro Leu Thu Thu Leu Ash Ash Cly Phe Gly 50 60
- The Gay Trp, And the Service The The Leu Asp the Lys The Leu The 65 70 75 80
- Phe Ser Arg Ala Asm Gly Glu Gln Phe Lys Cys Lys Pro Leu Pro Pro 85 90 95
- han han hap beu Ser Phe Dys hap Dys Dau Dys Asp beu hrg 100 105 110
- Val Tyr Lys Lou Asp Sox Asn Thr Phe Tyr Val Tyr Asn Lys Asn Gly 115 120 125
- The He Glu He Leu Lys Arg He Gly Ser Ser Asp the Ala Lys Thy 130 135 140
- Val Ala Lou Glu Pho Pro Asp Gly Glu Ala Phe Asp لخب Ila Tyr Asn 145 - 150 - 155 - 160
- Ser Arg She Ala Leu Ser Glu Ile Lye Tyr Arg Val Thr Gly Lye Thr 165 170 U75
- Tyr Leu Lys Leu Asu Tyr Ser Gly Asu Asu Cys Thr Ser Val Glu Tyr 180 185 190
- Pro Asp Asp Asp Asp Ile Sur Ala Lys Ile Ala Phe Asp Tym Arg Asm 195 200 205
- Amp Tyr Leu Ile Thr Val Thr Val Pro Tyr Amp Ala Ser Gly Pro Ile 210 215 220
- Amp Ser Ala Arg Phe Lym Met Thr Tyr Glr Thr Leu Lym Gly Val Phe 225 230 235 240
- Pro Val Ile Ser Thr Phe Arg Thr Pro Thr Cly Tyr Val Glu Leu Val 245 250 255
- Ser Tyr Lys Glu Ash Gly His Lys Val Thr Asp Thr Glu Tyr He Pro 260 265 270
- Tyr Ala Ala Leu Tur Ile Gin Pro Gly Asn Gly Gin Pro Ala Val 275 280 285
- Ser Lys Sur Tyr Glu Tyr Ser Ser Val His Asn Phe Leu Gly Tyr Ser 290 295 300
- Ser Gly Arg Thr Ser Phe Amp Ser Ser Gln Arap Asa lea Tyr Lea Val 305 310 315 320
- Thr Gly Mas Tyr Thr Tyr Ser Ser Ile Glu Arg Val. Leu Asp Gly Gln 325 330 335
- Ser Val Val Ser Val Ile Glu Arg Val Phe Ash Lys Phe His Lau Met 340 345 350
- Thur Lym Glu Ale Lym Thur Glu Asp, Asp, Lym Arg, 1le Thur Thur Glu Ile 355 360 365
- The Tyr Asn Glu Asp Leu See Lys Ser Phe Ser Glu Gln Pro Glo Asn 370 375 380

100a 3 8 5	Gln	СЪ	Pro	See.	Л гд 390	Val	Leu	Tiu	Arg	Tyr 395	Tive	Asq)	Ile	Gln	Tra: 400
Asn	1tm	Ser	Arg	Glu 405	Glu	Tor	Val	Asta	110 410		Ser	Авр	æp	Ттр 415	
Agan	Th:	ृष्ट्या	Leu 420	Ile	Tìur	Glu	Tru	Ser 425	ΕĴΆ	Ile	Gln	Lys	Glu 4 30	Tyr	Val
Ίγτ	Tyr	Pro 435	Val	Aso	Gly	Glu	Gly 440	Asn	Ser	Cys	Pro	Ala 445	AEP	Pro	Leu
ΘĴÄ	Phe 450	Ser	Arts	Ehe	Leu	Ly8 455	Ser	Val	Thr	Glu	Lys 460	GJY	Ser	Pro	As 1>
860					470					475			Тут		480
Pine	Pro	Tiur	Phe	ቸስድ 485	Gly	ALA	Туг	Veit	I.թթ 490	Olu	lyr	Val	Sear	Lye 495	Va).
			500					505					Тул 510		
		212					520					525	Thu		
	530					535					540		туп		
262					550					555			Gly		560
				565					570				GJD	575	
			280					585					λ ер 590		
		כעני					600					605	Arg		
	ØT0					615					620		Asn		
023					630					635			Arg		640
				545					650				Gln	655	•
уъф	Gly	Ala	990 117p	Gly	T	Sor	Gly	11e 665	Тух	Gln	Gly	ייונט	Tyr 670	Ang	Liye
Val	Leu	Ala 675	Arg	Gln	Tyr	ASQ	Va.L 680	Leu	Gly	Gln	Leu	Ser 685	Lye	Glu	Ile
	ьэц					695					700		Val		
Ala 705	Thr	Pro	Leu	Val	ፕት ድ 720	Thr	r)	Tìr	Tyr	լչ 715	Tyr	A SEP	Gly	Trp	Gly 720
Ascı	Leu	ፓሃ <u>ኮ</u>	Ser	Thr	Glu	ЛУŦ	Scr	ሊ ጥ	Оlу	Arg	Ile	ශීය	Leu	Glu	IJŧ

				725					730					735	
His	Pap	Pτο	11e 740	Tha	Arq	Thr	Ile	Thr 745	Gln	Сlу	Val	ГÀЗ	Gly 750	Æu	G)y
Met.	Leu	Asn 755	ī.le	Gla	Glin	Asn	Azm 760)	Phe	Glu	GJI	Pro	Ala 765	3eor	lle	Lye
Ala	Val 770	zy r	Pro	Asp	Gly	Thir 775	He	Tye	Ser	Thu:	Arg 780	Time	Тут	Arņ	ΊУΤ
Asp 785	Gly	Etvo	Gly	Arg	Tha 790	Val	Thr	Glπ	Than	леўі 795	Ala	Glu	Gly	}tie	800 УЈЧ
Thu	Gľu	Ile	Gly	Тут 805	ЪФ	Val	Phe	Asp	Лгу 810	I]ø	Val	Lye	Lye	Thr 815	Lou
Pro	Asp	СĴУ	Thr 820	Ile	Len	Glu	Ser	Ala 825	Tyr	aLK	Ser	Phe	Ser 830	Жis	Glu
Glu	Lau	11e 835	Ser	Ala	Ten	Asn	Val 840	Asn	Gly	ΉΈ	CIL	Leu 845	Gly	γја	Lou
Val	TYT 650		Gly	Leu	GJA	Arg B55	Val	Lle	Sex	Asp	Tha r 860		Gly	Gly	Arg
Lys 665		ദിധ	Tyr	Leu	Tyr 870	Gly	Pxo	Gln	GΙΛ	Азр 875		Pro	Пе	Glu	Ser 880
Ile	Thr	Pro	Ser	и16 805	Asn	Lys	Gln	Aem	Me t 890		Тут	. Yen	Тут	Tyr 89 5	
Ġĵу	Ser	Val	Ne t 900		Lys	Ehe	Thr	Thr 905		Thr	P.Sp	Gln	Gl n 910		Phe
Arg	Тут	His 915		Liya	Time	αly	ਬੀਮ: 920		Leu	. Ser	· Ala	Ser 925		Gly	Val
Ser	930 930		A ZO	Тух	Ser'	Тут 935		Pro	Set	Gly	740 940		Gln	Arg	(GJ v
Ser 945		Leu	Arg	ASA)	950		· Pro	. IJ∈	. Ser	Ser 955		r Gle	ያ ት ፕ	Len	Tyr 960
'Ita	: 19 ∈t	. Sen	Gly	965	lle	Gln	Arg	His	970	_	Ser	Pha	Gly	(119 975	
)៤ខ	: Val	Tyr	980		: Asp	Ala	Glin	Gly 985		Leo	val	. Lys	गाः 990		Glr.
ASO	Ala	Gln 999	_	Ala	r i ta	Ehe	: Glu 1000	-	Ası) AST	ı Val	. Gly 1005	_	le:	ıle
Tha	The 1010		The	T.ye	Asp	Thi 1015		Se1	Leo	ser	1020		VS)	Tin	Lys
11e 025		l Tyn	· Acq	> Ala	1030		Ar g	(Glu	ı Ile	÷ L∂⁄s 1035		j Sei	C Len,	Ile	9 S€1 104(
Ast	e left ne	. Sei	1 1 e	Glr 1049	vai 5	340	: Tha	Le	. <i>S</i> er 1050		r T IN	Lyz	ABT	1055	
Ile	e Seo	c Glr	1 A rg		e Tha	Se1	: £le	Аы 1069		y Val	l Val	L 196=1	Lyt 1070		3 Glı

- Arg Tyr Gln Tyr Asp Asn Asn Gln Arg Leu Ser Gln Tyr Gln Cys Glu 1075 1080 1085
- Gly Giu Gln Ser Pro Ile Asp His Thr Gly Arg Val Leu Asn Gln Gln 1090 1095 1100
- Ile Tyr His Tyr Asp Gln Trp Gly Asm Ile Lys Arg Leu Asp Asm Thr 105 1110 1115 1120
- Tyr Arg Amp Gly Lym Glu Tur Val Amp Tyr Ris Phe Ser Glm Ala Amp 1125 1130 1135
- Pro Thr Glm Leu Ile Arg Ile Thr Ser Asp Lys Glm Glm Ile Glm Les, 1140 1145 1150
- Set Tyr Asp Ale Asn Oly Asn Leu Thr Ang Asp Glu Lys Gly Gln Thr 1155 1160 1155
- Len Ile Tyr Asp Gln Asn Asn Arg Len Val Gln Val Lys Asp Arg Len 1170 1175 1180
- Cly Asn Leo Val Cys Ser Tyr Gln Tyr Asp Ala Leo Asn Lys Leo Thr 185 1190 1195 1200
- Ala Gln Val Leu Ala Asn Gly Tor Val Asn Arg Glo His Tyr Ala Ser 1205 1210 1215
- Gly Lye Wal Thr Agn Ile Gln Leo Gly Asp Glo Ala Ile Thr Trp Leo 1220 1225 1230
- Ser Ser Amp bys Gln Amp lie Gly Ria Gln Ser Ala Lys Amn Gly Gln 1235 1240 1245
- Sor Val Tyr Tyr Gln Tyr Gly Ile Asp Ris Aso Ser Thr Val Ile Ala 1250 1260
- Ser Glin Asm Gliu Asm Gliu Leu Met. Alsa Leu Ser Tyr Thir Pro Tyr Gly 265 1270 1275 1280
- The Arm Ser Leu Ile Ser Ser Low Pro Gly Leu Asu Gly Ala Glu Val 1285 1290 1295
- Asp Pro Val Thr Gly Trp Tyr Fhe Leu Gly Ash Gly Tyr Arg Val Phe 1300~1300~1305~1310
- Asn Pro Val Leu Met Arg the His Ser Pro Asp Ser Trp Ser Pro Phe 1315 1320 1325
- Gly Arg Gly Gly The Aso Pro Tyr Thr Tyr Cys Gln Gly Amp Pro Ile 1330 1335 1340
- Asn Arg Ile Amp Leu Am Oly His Leu Sar Ala Gly Gly Ile Leu Gly 345 1350 1355 1360
- Ile Val Leu Gly Ala Ile Gly Ile Val Gly Ile Val Ser Leu Gly 1365 1370 1375
- Ala Gly Ala Ala Ile Ser Ala Gly Leu Ile Ala Ala Gly Gly Ala Leu 1380 1385 1390
- Gly Ala Ile Ala Ser Thr Ser Ala Leu Ala Val Thr Ala Thr Val Ile 1395 1400 1405
- Gly Leu Ala Ala Asp Ser Ile Gly Ile Ala Ser Ala Ala Leu Ser Glu 1410 1415 1420

```
lys Asp Pro Lys Thr Ser Gly Ile Leo Asp Tro Ile Sor Ala Gly Leo
Gly Val Leu Ser Pho Gly Ile Ser Ala Ile Thr Phe Thr Ser Sor Leu
               1445
                                   1450
Val Lys Ser Ala Arg Ser Gly Ser Gln Ala Val Ser Ala Gly Val Ile
Gly Ser Val Pro Leo Glo Phe Gly Glo Val Ala Ser Arg Ser Ker Arg
                          1480
Arg Tro Asp The Ala Leu Ser Ser The Ser Leu Gly Ala Asm Ala Ala
                       1495
                                           1500
Ser Len Ser Thr Gly Ile Ala Ala Ala Ala Val Ala Amp Ser Asn Ala
                   1510
Aso Ala Ala Asat Ile Lou Gly Top Val Ser Five Gly Five Gly Ala Val
                                  1530
Ser Thr Thr Ser Cly lle lle Glu Leo Thr Arg Thr Ala Tyr Ala val
           1540
                               1545
                                                   1550
Aso His Glo The Top Glu Leu Ser Ser Ser Ala Gly Yor Ser Glu Glu
                          1560
Val iya Pro Ile Ary Cys Lou Val Ser His Ary Trp Asn Gin iya Gin
                       1575
<210> 7
<211> 18
<212> DNA
<213> Artificial Seguence
<220>
<223> Description of Artificial Sequence:oligonucleotide
<400× 7
acacagcagg ttcgtcag
                                                                  18
<210> 8
<211> 18
<212> TBAA
<213> Artificial Sequence
<223> Description of Artificial Sequence:oligonucleotide
<400> 8
gycegaages ctmaachd
                                                                  18
-210⊳ 9
<213> 20
<212> DNA
<213> Artificial Sequence
<220b
<223> Description of Artificial Sequence:oligonurlectide
<400> S
```

```
attgatagra ogeggegese
                                                                   20
<210 - 10
<211.5 22
<212> DNA
<213> Artificial Sequence
<220>
<223> Description of Artificial Sequence:oligonucleotide
<400> 10
tiginacqiq qaqeeqaset qg
                                                                   22
<210> 11
<211> 37948
<212> 1MA
<213> Photorhabdus luminescene
<220×
<221> CDS
<222> {35171)..(18035)
<223> orf5
<220×
<221> CDS
<222> (23768)..(31336)
<2235 hph2
<220>
<221> CDS
<222> (31393)...(35838)
<223> onf2
<400> 11
integrings conjugaçan tatgertate groupitaga acquiscan etcaatgang 60
ccaagatgtg gtatatgcas graetgcatc tollaggega tamacetest ctategites 120
gitcagagig gagcaaarog agiitaggog gogcigoogg aanggagaga caaaagcaar 190
acgrocasgo aatggoogot, obyogsoway gtystyttag toggoaraac aaccogsoay 240
atottitott goomnaggid matgaagigs igosaaacia tiggaaaaaa iiggaacaan 300
ggotgtataa ootgogtost escotoarta tigsrygoom accyctacat otgochalet 360
acgotacaco ggosgatoca sasagosttao ttagogoogo tytogotago teogaagyty 420
ggytagetet etcacageca titatytoae tytyycytti erracaeaby etgyaaaacy 480
egogbygtat gytcagtcag etcsetczał tegyetetec getaczaaat attategaac 540
grasgatus gymagettta ascaegetet tyrageates agrageages etgatattys 600
cteatetemy cutacoggae азаассятся сададсідда ідсуувалля абоділетра 660
альнатожнь насодуруст слителодот страсадств свазаведств (исдествой 720
statowales gygligaaasu ogggetatag sattgestee eteegitiget geerkunges 700
rigocotycs agostosogi otygogygog otygogothya totygogoco aaraittiog 840
głakusztow isposytago cyttysysky cymitymyk wychacagyt aatyliatyg 900
```

estinicogo cagigigaty aacamogasy mygataaaat cagmmagulua gaayoolalo 960 ggggtogeeg teaggaatgg gaaatocage gtaatestge ogsagesgag ataasacaga 1.020 togatectea acticastea etggeagtse geogtgaage egesystatte eagamageea 1080 quitament commande expectate decastige titrotype ogtamenta 1140 gtastcaage gttgtactae tggetaegeg gteggk:tagu tgetatttae ttecaatttt 1200 acquittings equagegest typictgatgy objeatings that within galacticate 1260 agacogogo nagotttate naaccoggog eetpgcaggg aaccomatgeg gg#b\$antgg 1320 ctogtesaso ct/gateoby autotogogo anategraga teconstitis aggingente 1380 ascgryptick gyaegtogaa oggaccatti cattymugca acactaigga gcactyccag 1440 assautogut tastutayoo acaaggattt etacootgot agraggtggt acaambgsct 1.500 creatigatga testroxytt scattagass acquecaset tagigecasa steicicigi 1560 caputetyte attagataat gaetaeeegg alagemaegg ogtoggeaac attogaegea 1620 transparat captgreace troccapeer tottaggam; atateggan gtacaageta 1680 ttotgtocta oggaggaagt gaaateggat tagetgaaag etgtamatem etggogatet 1740 ctoatgrout caatgaragt getcaatter agtiggatit (aaccaige) aagticeige 1800 cytthywwgg guttgogatt aacgatacty gomeathyse autozytthe coccaatgog 1860 actytoware amgazzacat gitgesgart tigagigais tiaticipea tattogetat 1920 scratcogem ssiawoooc togattaaat accaaaaaca ggcteciaaa egggeotga 1980 actiticacy aatatatano actoacayte tyetetetti acciyiciya eyeinyitab 2040 aacagagata titectitic icqtqqqqqq qqlquoxtac tataaaatai caacccicti 2100 ettittrata atatgeaata lytaacaat yeaattatti eattiagita tigitaacta 2160 gttatattac tiatgetgie attobament bigitatige atcacaatag contiteset 2220 aaataataac gttgtgaaat agtbgatagt tzaatggtgt tittettiag cogitatlit 2280 caacocaatt teagacegek akkagaegit accigigitg ceittyitti galegalata 2340estaspotta titatatoos ogytaniosg arregostas atgititati tarciascai 2400 ttaanngan tanacatgaa cacactcaan trogantaty nunncypytt aytagcagyt 2460 titaataate taacegatat ligiombete teithigmog aachtegeas aaaagigaag 2520 gacasactot calggtonca gaccossage tistatotitg asgonosyca agigonosing 2580 gaceak/Urn tycatgaage coatationig aaacgegoda electroatit acazagigog 2640 guncalciti cootgaccae accidatget gaccagesag gitalcastag capatitype 2700 astrogorus guasstatze agreecezge gesatttett eestgiitu teetgegget 2760 tatttagetg aactttatog toxogcmogg patriacatg sigasaatte tatttatest 2820

tiggatacac poegicogga icizaazioa tiggipetea gecagaazza taiggatacg 2880
sagattteen esetttetet stetaatsse ätyttsetag aggstattas aactetsete 2940
aaggacaage tgetggugge tetgnagmat attamatete tgtggaagga egagetgetg 3000
gaggettiga agaztattaa acetetytee aaggaogate tyelgoogge titgaagaat 3060
attamaeete tytoemagya oyatetyety gayyettiga myantattam mertetytee 3120
aaggmogate tgetggmage titgmagmat attamaeete tgiecmaggm ogateigeig 3180
gaggethtga agaabatkus acutetgtee aaggaegate tgetggagge httgaagaat 3240
aktaaacete Egiccaagga egateigeig gaggeliiga agaatattaa accimigioo 3300
aggregate tacaggrafy tettgesatt ctaticaate tggacagera cactosagta 3360
atgassgret taterastit regestitet gewatgatge catateacya beettatesa 3420
agostgosta aggitistica attacaggot complisting aacaostis. tastacatca 3480
ctagaaacqa ctatcgatga actamamtat campettett tgttggammi: tamttettet 3540
giclegocia ambiatitae intelligaci ganganatia etamanican igraegengi 3600
ototetemay wasattoteg teatattama octoectaa teggammanoc ggammatatoty 3660
esaagttatt sceeksigag tgatgaagag titagagatt tmektamaat aagaactata 3720
cttcttccag aageageset ogcombtact patettgcat cycytactoc tagtacacaa 3780
cagactatom assatorings Elektomient etailmassa tisatasett tettomieta 3840
ticaaageta taaacttate acegacygta klaagtgyna teeteeyeag catcagozoa 3900
gazttoaata tosataaaga satattacaa saaatohtto gigitaasia chafatgosa 3960
ogttætggta tigæcastga gastgsatta atastatgsa mogtassaat tisassatai. 4020
slemalgery generotes, tongttigat ogtitatita elicococaa actgaeiggt. 4080
compatitit contomating tactomizant attratition coctangeng taccascence \$140
tyyaataaaa cagtacttaa acytgetttt sacctcgatg atatotestt saatcgacta 4200
ctamematta committeecquit caminates gammatta ctaminatat agammatett 4260
Extratistic staggaczae attactagra gatatorate agttaectat tgatgaertg 4320
gggttactet tegsasecuat aggtamagga acaaccastt taleteggat tacteetgse 4380
antotypita etetanitus emaartetai geigieneta geiggeiang iangemanag 4440
tygagtgict alcagitgit talgetged, edugatment alaacaasac colaeccog 4500
gaastarggs atttactggs twoogtotec eatggetige sagatifies teageogaig 4560
ttgaaagetg aagaagateL нужуюжных: малаждыаат tgeagegkgu смаружалат 4620
стуравават тосгудавае созденесь сілоную доздравава вустиную 4680
agnotymata angotyaaga gacccaogaa aaagccysga aaaacctaga tgaggtcagg 4740
assautotgo cazzagotat ateteettat ategergoog etetgeaatt accatetgaa 4800

cabgroggest attocatant catchgoyes gatastring asucceptost aggrassastg 4860 acageggasa aattatggaa etggttgegg asaasteeeg tüsuggetez aeetgaatte 4920 casaaacaag ctgaacctgt ggtccagtat tgccagcgcc bggcaccact agcgttgatt 4980 taccytteta cogyectias cyassaesee tiasyteiyi Utybyacana gecycasese 5040 tttgttatta aaaocaaago aocegaaaca actgaaacaa caccagoaca tgaogtatca 5100 acmetaatyt exctaacycy tittactyse tyyyttäset eactagykya ammogeetet 5160 hotgtactwa cogsatttaa aasaggaaca ttaatagcag aactattggc (waggotatg 5220 aatritigata aaaaleteet gaagemagee magattoagg caasaactgs titigtoomae 5280 typocatota togaczacot attyczytyg attaacstot cacytoaktt gazcatotot 5340 conceangeen bittoczonet gastonagta tigacogoag aassitocops tambiataes 5400 coateggass segongotgo gatattaaco googgactgg acaeccasas gactaacaco 5060 ctacatgogt ttotggetga glutogoagt gotgogttaa goarstacts (attitattot 5520 cataancaas aagatogaga ayosegsaas (atacagtaa timaaasoog tgsbgutota 5580 tatosatace tattgatega taaccasgtt tempcogeca ttaaasetac agagstoget 5640 gaagetatog etagtateca actgtatatt aacryggggt tgaasaatat ggagggagat 5700 actytomoma gtytomocay coyotomite ticsocaect gygatamata castaeecyc 5760 tacagosott gygocogitat gypotaaacte etittactale cagagaatta estegateny 5820 acyctacy(a bbygycagae znasatgatg gatacyttyc tycaatccat cagecaaage 5880 castressta togataccel; appagateco titazatett acetaacstu attogascae 5940 gtggstaate tggaaateet sagegookan oatgasaasa ttaataatga temaggatta 6000 acttacttta teggacgtag taasacegss gbyaatcaat attattggeg cagtgtggat 8060 caraataaat ocapoqaagg taaalikooco gotaatgoot ggagtgagty gtacaasatt 6120 gattgtecaa ttaaccecta casagatact attcaccegg taatttteca atctegootg 6180 tatottatot ggotggaaca aaassayyey selasakagg aaggtgataa sekugootog 6240 ggttalkátt algesotgsa attagogsat atcogttaig elegozosig gaatasaku 6300gtrarcutte augumanca massatatro gattiasatr teggammatas macacotega 6360 etttaetget caagetttea aggeagagat gaattgetgg tgatgtttta taaaaaacaa 6420 qatcasttas alcastacae associata eesstasaag gackatatat caetteesat 6480. etglolitoka egymaatgad acetgaaaat caraaaeets argottataa acagittigat 6540. antaanagus khaktggtgt cantaatogo tatgoagaaa gotaogaaat coottoatoa 6600 qtaastaqta ataacggtta tgattyggga gatggctatc tgagtatggk ytatggcgga 6660. aatatttoag coatoaaact ggagtootos toagstaagt taaaactoto soosaggilla 6720

aganttatte	afeetgyact	tgtaggeega	caacgcazee	autgesacet	gatqaagaaa	6780
tacogotcape	ttggtgataa	attattatt	tatactactc	taggtattaa	ccccaataat	6840
ttgtogaata	esesticat	ttaccctgtt	tatcogtata	gtgggaacac	taccestaat	6900
gregadaggac	gtetgetgtt	ttategagaa	egtactacta	actttgtaag	ega atgg tta	6960
cetaacettr	cetetggete	tcaagaaatg	tocacaaoca	ctggoggtga	cattagtggt	7020
exchallggct.	atattqataa	casalcatagt	gaogatgitic	catttaaaca	atatttctat	7080
atggatgace	acygtyytat	tgacactgal:	gt.tbcsggga	tattatetat	taataogaac	7140
attaatcatt	casesyttes	agtaategtg	вандарданц	gtatoacaga	gowaactttt	7200
gtagegageg	aadecagliae	lighocomscc	aatoogtoox:	gabbagaaga	aatgaattat	7260
cagtttamag	agottgaaat.	agstatetee	acactgacs);	ticatestaa	tgaagcaagt	7320
alitgatal/#	octttstege	atttgetgag	aaatttgacy	etaataguaa	tgatogtasc	7380
(Jaggggeeg	aacatttcag	tattogtatt	atcasaasag	oggazactge	Lastgoodtg	7440
acectgrace	ataatgcaaa	ogggggggcaa	tatatgcagt	ggggaaactc	ttgtattog:	7500
cttaataogo	tatttgeeg	tomattaatt	agcogagoca	acgeggggat	agetectett	75 6 0
ttgagratgg	acaetcagna	tatteaggaa	cetazattag	gagaagattc	tootgetget.	7620
atggaaccaa	tggacttcaa	ордодоськае	agoststatt	totgggaact	gttctactec	7680
accccgatge	tgattgetea	acotttocto	CHOQUACONS	acttogatga	ggctaaccgt	7740
tggotgaaat	abgitaligges	unuatouggt	tatattetca	abggtcaaat	geaacattac	7 800
ാവുമുള്ള	trogecests	асиндаррас	ectagtilgga	acgatgatec	gttagattca	7860
tttgetcctg	ataccatage	tcaacatget	ocaatgcact	acesagtugo	cacctttatg	7920
egeacectag	atctyttgat	сувасуууда	gattaogect	atogocaatt	99090999	7980
acactegetg	aagccaaaat	gtggtatatg	caggeactge:	atctattegg	tgataaacet	8040
matctaccac	tcagttcagc	atiggeatigat.	ccagagntag	88 9899 0090	agetettgaa	8100
asscaecsy	cacatgccs	egaaatagraa	gatttacgac	aaggacttee	tacateesca	8160
gggtctaaag	atymaatosa	aac agatett	ttoctgooge	eaglingerigg	agtgatgetg	8 220
agetactgge	agravactaga	acaaoggttg	tataacetge	gccataacct	Ctctattgat	6280
ggtcasectt	tacatttgcc	tattttcgca	асассвосью	atomasaasc	getgeteage	B340
geogotytag	coagttes(a	aggtggaagt	astoticcat	cagoatttat	atcagtgtgg	B400
ogtiticatio	etalgotyga	azialogeologit	agtatggtca	gtcagetaae	ceasttogg:	6460
tomacattqu	amoutattat	ogosegtesa	gatgeggagg	cattaaacac	gctgttgcaa	8520
antoaggdgy	cagaactgat	attgaccaat	ctragcatac	аддагаалас	catccsagag	85B0
ctogatocky	а авыза нууд	gotagaaaan	escondeducid	gaacccagte	gegttttgat	8640
agetacagus	aattotaoga	tgaagacato	9803GGGGGINB	аахаасадде	aatggegtiq	B700

cytycttog tugctgycat ctctacagoo cttcaagoat cacatotgge ggyegragea 8760 ctigatotym ogoddaacet cticyycite grimatymig geageegity gygygemetr H820 geoceaageea eaggtautgi; eatggagtte teegeeagtg thatgaacse egsageggs (, 88(ii) aaaatoagoo aatutyaago ciisoogtogg ogtogtoagg aatgggaaat icagcoblast: 1940 ascreegings capagerigas acasastegat metesacttm gittemetage agrigoryemyr 9000 gmagecegoug tettgoagaa aaccagoota asaacccaac aagagoagac toatgoacsa 9060 stassettee tycaacetaa ytteaytaat cayyeyetel teaactyyet geytyytega 9120 lughorymra ittactions giicinigat itsaogytag etopetgiii gaigyrygaa 9180 alguntate getgggagae taacquatace quatoscqut tratessace eggegeetgg 9240 саууувасос atgooggttt potogogggt gannouttam tgotgaatet ggcacagaty 9300 gaagatgeec acctgmaaca yymtaaanyn gtantyyagy tagmacytac cyttteyety 9360 googaagtot atgoomaatt accepsaagat aaatttatoo tgactoagga aatagagaag 9420 ttogtgagta seggittcagg cagggreggr aaggecaata ataagcigge gittagtace 9480 astaccanis octobetaga agogtocatt togttatota octogaacsi: tagoagogat 9540 harrongsti otsinggiaa aacoognogi attaaacaga toagogitso octgorogio 9600 etgetaygac eetateagga tgtgemagem attetgtett acageggaam agestekynn 9660 tippotgaaa giigomaato attagoggit butcabggga igamigacag oggicegiic 9720 caactgyatt tosaogskyp (asabhnety erytteyasy gaatosaant egatgaaggt 9780 augstypenat typgistions aastyssatt aytsaagsay assesseege cynaasagge 9840 Asacsanaag coatgetyga gagtetgaan gamatoatto tyostattoy otamasoatt 9900 ogcomataac gattitaatt aagtgotama acaggoocot aageggggoo tgemaggagt 9960 ctttcatgcs aaattcacaa gatttcagta ttaragaact atcattgccc ammggaggag 10020 gegetaleae gggmatgggg gaagetttaa cereesergg gerggatggg atggregege 10080 tytetetyee yttyeetate tetyeeyyye yryyttatye teeyteaete yeettammet 10140 acascagong ogsengtame agescartin grothnysty gnattysame gutabysowa 10200 topgeogoog cacecattit ggegtteese attatgatga aacogstacc titetgggge 10260 cognitigate getectigate etagragate astrongoga constigues timengoga 10320 towarttage carrecettt accepttarreg galacogtte cogtetegag agteatttea 10380 geogratiggs statiggess eccaeggess exercisegue saciggesse acagaithti 10440 ggutgatela (agguragat ggarsagter stitaciggg taaatoxoox coaggoogga 10500 Reagnasce: gtosganate actessacag caesatgytt getagasge: tebytytese 10560 cucatgytga acaaatitat tatcaatatc gyyccgagga taacaccgyt tgcgaagctg 10620

atyaaattac	totocation	cadilicalcad	ogcaacgtta	tetaescaes	gtgtattarg	10680
ÖGANKKAÄÄRK	водгосводгава	acgttacocg	gtctggatgg	CSGCGCCCCC	CCDCSSGCSG	10740
ac:tggttatt	ctatctggts	tttgattang	ಡದಡಿಕಾರಡಿದ್ದಾರ	taaraacetg	премодооде	10B00
cagnatitite	gactacaggt	agetggettt.	atroaccaria#	cogtttttcc	ogitatgaat	10860
atgyttttga	gattogtacc	egeegettat	geogteaggt	stigatykat	naccacetge	10920
aagetetgga	tagogagata	<i>coagractica</i>	acggaccaac	getggtttca	ത്യാന്ത്യപ്രമാ	709 80
tceattatga	വുള്ളുകളുന്നുവര	atugocagoa	cyclygtatt	cyttegtoga	gtaggocacg	11040
ayodagacyg	tectgrogic	accettgextge	cattagaatt	ggogtateag	qatttttcac	11100
rgotacahaa	aga ta gat <u>o</u> g	caatogetga	atgtgetgge	aaacttcaat	geeatteage	11160
go tgg caact;	egtigalota	8689 309 089	gatteeegg	totgotatat	caagstaaag	11220
gagaatggtg	gtaccyctcc	goscaacget	Cogodinaat	togetcəgat	googteactt	11260
ggge <u>aaas</u> at	gcaacctttg	toggttatoc	ettmelitgea	aagtaatgee	togatggtgg	11340
atatoaatgg	agaoggecsa	cttgactggg	ttatcaccyy	ac re ggatte	oggggatate	11400
atagtcagoa	tecagatoge	agttggaæec	gttttaccc;;	gottowaeget	ctgccagtgg	11460
aatatactco	tocaegogeg	caactegeeg	atttmatggg	agotggaett	totgatttag	11520
tactgatogg	coctangagt	gtaogtttat	atgecastae	CCECGARAGE	tttgccaaag	11580
gmaaagatgt	agtgcaatce	ggtgatatca	cartgecagt	accogggwycc	gatocytyta	11640
saghtaggi:ggc	atttaytgat	gtattgggtt	coggtcagge	acetotygtt	ධ්පමේටාට්ප්ට්රුවි	11700
ogsctasayt	czectycty	witaatotgg	diamodianoi	ttttggtcaa	ccaattactc	11760
ttoogygatt	tagecsaces	gesgc <u>s</u> acgt	Utantoctgo	toagittai	ctggccgatc	11820
tagatggeog	oggeoogaet	gatotgattt	atgittacac	agatogtetg	gatatettee	11880
tgaataaaag	්රීඩයාපයා ල්ය	ttogoogeec	cogtaactet	coectteres	gccggagtgc	11940
gttttgatca	tacctgtcag	ttaceagtgg	omgat gta ca	agggttaggc	gtogecages	12000
tgatattaag	tgtgoogoat	atgactecce	atcactggcg	t.tg og atetg	accaaracaa	12060
eaccytygtt.	acticagtigaa	абдаасааса	atatgggggc	tcatcacac;	ctgogttacc	12120
ytagttccyc	ccagtictogg	otggatg aa a	sedecaroddc	actggatgcc	99acaaatau	12180
cagtttgtta	tetaceette	coggtacace	ccctatggca	aacqqaaata	gaggatgaaa	12240
1.cagcggcaa	cessttagtc	acaatactac	gttatgcaca	tggegeetgg	gatygangiy	12300
а 9 09өджөі;L	togoggatti	gyttatyttg	аагадааада	cagocatosa	വൂർമിവാഷൻറി	085SC
grægtgogor	agaetigosica	concetegeae	tgacccaagg	caacgogoot	gaactcscat.	12420
caccogogot	gacornagge	eacoxtoxay	aactowcacc	acctgcpatg	ассеваадся	12480
ecgegoctga	acteacates	caagaggatga	උෆයුෂඉලදාස ස	ородосадаа	ttcacatcac	12540
cogrected	ccaeggceet	ўф ўссадад с	tuacaccace	tgegatgaee	aaaaactggt	12600

atgecacogg	aatacooatg	atagataxou	cettatogac	अप्रयक्षित्रस्था	catggtgatc	12660
accaagettt	tgooggtttt	teaceacyct	ttacgacetg	gcaagatggt	caagetatic	12720
tgetæææe	ggaaaatgat	aacagtcagt	actgyctaaa	cc89gcactg	ನಿನಚಿತ್ರಗಳು	12780
tgctacgcag	tspactstac	ප්රථන්ධ වාද්ධ	googtacaca	ggazaaestt	octacacag	12840
toactgaatt	togeceaeag	gtaogtoggt	tacagoatae	cgatagooge	tacettetee	32900
tttggtæte	totagttgaæ	agrocythaet	atcattacga	acqtatcgcc	agoga(coctr:	12960
aatgcagoca	amagatbacg	distimasm	sintaittee	tessoogeta	aaacaggttt	33020
oggtacegta	toceasporage	control	CARGUAGTOC	gtatoctgat	acqttgcctg	13080
ataagttati	tg ot aacago	tatgatgecc	А 9000000000 0	attaoggete	Bookatoaac	13140
agitoagitig	gnatka(idig	accescents	ccalletest	gttaggatta	coggatagta	13200
යයටයපවයවිද	tatetttget	tatagcgctg	aacstytexx:	tactogtggt	ctabelictgg	13260
aaateetaaa	tgataaaaat	agtotgattg	oggagaataa	acciliogt <u>o</u> ma	tacctoggoc	13320
DECOMMENS	cgtttatacc	980999988	atgczaogec	atogosseog	осидоверве	13380
aagogutgat	typoottoacc	gagacaacag	tatttaatoa	atocacecte	teagogtitg	13440
atoggaptat:	ctcatctgct	castigicaa	cgacgetgga	acaagroyyn	taccagcaaa	13500
cagattetct	<u>etterry</u> cyc	act ggagaa g	ateaagtetg	ggeagetegt:	ogtøgetata	23560
ctgattacgg	cecedocae	cegbtobggc	99009оваал	acedadcaec	Actoazeten	13620
ogggcsasat	caegeteaet	tgggmtgcma	actat.hg cg t	cytca cacaa	accogggatg	13680
oggetggaet	gaczacetea	gocagatatg	attggggtti:	tetgaccece	gttcaactma	13740
oggatatom.	ognosatoag	caccttacca	cyctgyaty:	800 9990 00	ರಾಹತಿಕೆಯಾಗಕಭ	13800
tgogottttg	GGTTPIACCGEAA	aacggtaaga	tgactggttva	ttetteareg	gaeseestal.	33860
cgttttetee	accatetgat	gttgeogoog	cgattaa gkt	ತೂರಮಾರಭಾರತ	atocolopias	13 5 20
cacagtgica	qqtetaegea	occessanget	ggatgcccat	attaaagaaa	accetraata	13980
acctggcaga	acaa dagedd	aaagagttat	atascacxxy	aatoxtoacc	gaagengge;	1.4040
goztotytac	cetageteae	caccactooa	taaaaagoon	saytocagto	accongcosa	14100
tosatetete	anaeggeagt.	ceccgtttae	contratag	cotoscattg	actacygatic	14160
gttatgaccg	cgatettaag	caec egettc	gtcaacaagt	agtattcagt	galggelttg	14220
geogtttact	graagratet	gtacgacaty	вродцеорая	agretgycka	cytancenag	14280
acggegetet	ggtgaceann	والمهيدوني	схавалодед	atgggaggtt;	кордумерса	14340
ctqsatatga	caataaggga	caaccyatar;	goacctatca	accctatttc	ctcaacgact	14400
gycaataogt	cagtaatgac	agtgr e nggr	уунсадажда.	agoctatgca	gatavocatg	14460
licialgatee	cattggtcga	g asstrskgg	licactacege	aaaaggct,gg	ttoogtogaa	14520

octtett.car; toottegettt actetoaate aagateaaaa teacacagut acteaggitea 14580
σ уусландан денадентут нандендуты надендутын налидият τ
ogcomystty saggyogga aacataacat aatatagagy tganaogtyt cattcataat 14700
gungknagek actmaartia tgayttigitt gatcaltigit titattigogg cetigigigigi 14760
ettagtaagg tacctcattg atatacaaas casacsalgt aantogsatt ggatcaacgt 14820
actetytema etestiatet eeigititae egytatelika aangaetye tyagititiga 14880
aagoggegge agoooctata temetittee gatteeogge etattiggen ecnoggmaag 14940
ttotggattg azotggatet pgogtogret ttttatgeat tatogegatg atggaggaaa 15000
greataagye attoccactg oogeaaaaac estetytete oggosyttaa acogggaast 1.5060
taootartac aactattgta agaaaacgaa tatatagaaa aactazcatg cagetaassa 1,5120
etgogattge agasengstg scaracaacg occeascaar gaggtamate atg ass 15176 Met lys 1
asc atm get cot ass cit tab cas asg acc cot gio gio asc atc tac 15224 Asm Ile Amp Pro Lys Leu Tyr Glm Lym Thr Pro Val Val Amm Ile Tyr 5 10 15
gst sac ogs ggt cta sog ato ogt sac ato gsc tit cac ogt acc acc 15272 Asp Asm Ary Gly Leu Thr Ile Arg Asm Ile Asp Rha His Arg Thr Thr 20 25 30
gca aac ggc gat acc gat atc cgt akt act cgc cat caa tat gac too 15320 Ala Aan Gly Asp The Asp De Arg De Thr Arg His Gln Tyr Asp Ser 35 40 45 50
ctt ggg can cha agu caa agu acc gat cug cgt cha tat goa guc aaa — 15368 Leu Gly His Leu Ser Gln Ser Thr Asp Pro Arg Leu Tyr Glu Ala Lys 55 — 60 — 65
can mass tot amo tit ofto tigg can that got tity made ofto asi: att tig 15416 Gin Lys Ser Asin Pho Lau Trp Gin Tyr Amp Lau Thr Gly Ami Ila Lau 70 75 80
tgt aca gaz age gte gat get ggt ege act gte act ttg aat gat att 15464 Cys Thr Glu Ser Val Asp Ala Gly Arg Thr Val Thr Lou Asn Asp Ile 85 90 95
gam one out con eta etg ama eta aut gon aon ent etc ata esa aux 15512 Glu Gly Arg Pro ian Leu Thr Val Thr Ala Thr Gly Val Ilo Glu Thr 100 105 110
ega caa tat gas ang tot kee ota eee ggt egt etg ttg tek gtt ace 15560 Arg Gln Tyr Glu Thr Ser Ser Len Pro Gly Arg Leu Leu Ser Val Thr 115 120 125 130
gaa caa ats cms gea exe ace too ogt atc ace gaa ogd otg att tgg 15608 Glu Gln He Pro Glu bys Thr Ser Arg He Thr Glu Arg Leu He Trp 135 140 145
get gge aat age gea gea gag ama ame ent aat ett gee age emg tye 15656 Ala Gly Asa Ser Glu Ala Glu Lys Asa His Asa Leu Ala Ser Gla Cys 150 155 160
gtg oge cac tal gad ang gog gga gtd and oga tha gag egi: thig tow 15704



Val	МÜ	618 1 6 5	Тут	ASD	'lhr	Ala	Gly 170	Wal	Thr	AYJ	Leu	Glu)75	Ser	Leu	Ser	
							caa Cln									15752
							gat Asp									15800
Ctg Ctg	get Alæ	qat Asp	yad Yad	atc Ile 215	tac Tyr	aca Thr	acc Thr	ctg Leu	age Sec 220	gre Ala	cti. Phe	gat Aep	gcc Ala	acc Thr 225	eja arc	15848
get Ala	tta Leu	cto Leu	act Th r 230	C)D C)D	açıç Tiur	990 Двр	g¢g Ala	ава Lyв 235	Gly	aac Agu	att Ilo	egn eag	өдд Ахү 24 0	cta Leu	euce Time	15896
tat. Tyr	gst Asp	gtg Val 245	gee Ala	ggg Gly	Clin Cag	CEA Leu	aec Ae n 250	ejà aaa	agc Ser	tog Trp	tta I <i>e</i> pj	acc Thr 255	tta Leu	esa Lys	ges: Pab	15944
cas Oln	eeg Pro 260	ĠĮ≀i Ôæ9	Gln caa	gtg Val	att Ile	atc Ile 265	Arg Arg	tee Sor	ctg Leu	acc Thi	tat Tyr 270	toc Ser	M# gec	gee Aln	gga Gly	1,5992
caa Gln 275	aaa Lys	tta Leu	ogc Arg	gag Glu	даг Glv 280	cac Bis	Oly Ogc	aat Asn	ggt Gly	gtt Val 285	atc Ile	acc Thr	gaa Glu	tac Tyr	agt Ser 290	16040
tat Tyr	goa Glu	ccg Pro	gaa Glu	acc Thir 295	cæa Gln	cag Glu	ctt Iæu	atc Ile	ggt Gly 300	acc Thr	aaa Lys	ecc Thr	cac Ris	egt Arg 305	eeg Pro	1608A
tma Ser	gat Agp	ger: Alæ	AAG LVS 310	Va)	ttg Leu	caa Gln	дас Авр	cta Leu 315	Ogt Mry	tat Tyr	gaç Glu	tat Tyr	900 Asp 320	cog Pro	gta Val	16136
egà àise	aat Asn	gtc Val. 325	atc I)e	agt. S c r	atc Ile	egt Arg	ast Azn 330	9ac ABP	eca Ala	gaa Glu	gee Ala	acc The 335	oge Arg	isto Phe	tog Trp	16184
cac Ris	aat Asn 340	cet Gin	aaa Lys	gtg Val	geg Ala	erg Pro 345	gea Glu	aac Aen	act Thr	tat Tyx	acc Thr 350	tac Tyr	gac Asp	toc Ser	ttg Leu	16232
tat Tyr 355	Cag Glin	ctt Leu	ate Ile	Ser Ser	ಥ್ಷರಾ ಸ್ಕೃಹಿ 360	Ππ	Gly ggg	ege Arg	gag Gìu	atg Met 365	Vya Boll	aat Ago	ata Ile	ggt: Gly	cag Gln 370	16280
Gln Cea	egt. Ser	asc aen	caa Gln	ett Leu 375	ecc Prvi	tec Ser	ete Leu	acc Thr	cta Leu 370	ent. Pro	bot Ser	Jast) Jest)	aac Asn	aac Asn 385	acc Thr	16328
tac Tyr	acc The	yan	tat Tyr 390	Tìm	cgt Arg	act Thr	tal: Tyr	act ურლ 395	tat Tyr	gac Asp	egt. Arg	ejà ôôc	990 Gly 400	Asn	ttg Iæ:	16376
act Thr	688 688	atr 11e 405	Glr	cac His	agt Ser	toa Ser	eeg Pro 410	Ale	acg Thr	caa Cln	aac Aan	ക്കോ Aജന 415	tac Tyr	acc Thir	aca Thr	16424
aan Asn	ate Ile	acç Thr	Ve) gtt	tet Ser	aac Asn	ogg Arg	agc Ser	ast Asn	cy: Arg	gca Ala	gta Vel	ete Lou	age Ser	act. Thr	ಗಳ ಗ ಭದ	36472

WO 99/42589

620 425 430

	WAU					دسد										
acc Thr 435	gaa Clu	gat Asp	eeg Pzv	geg Ala	caa Gln 440	gta Vel	gat As p	get Ala	tta Leu	ttt Phe 445	gat Asp	gra Ala	Oly Oly	gyy àga	cet Ris 450	16520
Gìn Gìn	aec Aen	acg Thr	ttg Lev	ata Ile 455	tca Ser	gga Cly	caa Olm	aac Asn	ctq Iæu 460	aac Ası	tgg Trp	aat Aso	aca 'Jinc	agç Arg 465	ggt. Gly	76568
							gtg Va)									16616
							agt Ser 490									16664
atc Ile	ant Asn 500	gaa Glu	cag Gln	cag Gln	acc Thu	ago Ser 505	agc Ser	aac Aso	tet Ser	Gln Gln	aca Thr 510	cag Gln	aga Arg	ate 1le	act. Thr	1.673.2
							ogt Arg									16760
							aca Thr									16808
				. ATE					Gla					Āsp	aat Asn	16856
			Arg					Asn					.9er		ett Leu	16904
		λερ					He					Glu			Pro	16952
	Gly					TIE					Arro				agt Sar 610	17000
tat Tyr	aaa Lys	acc The	ato : Ilc	ogt Arg 615	Tyx	tcs Sen	ggt Gly	aaa lys	gag Glu 620	Arg	est Asp	goo Ala	acc The	gga Gly 625	cta Leu	17048
				Tyr					Pro					TIP :	tta Leu	17096
			Pro					Āæ					Tyr		; Met	17144
		ÀЮ					Leu					Gly			oon Pro	17192
	r Ile					Ala					Aer				уні. к Авр 690	17240

•

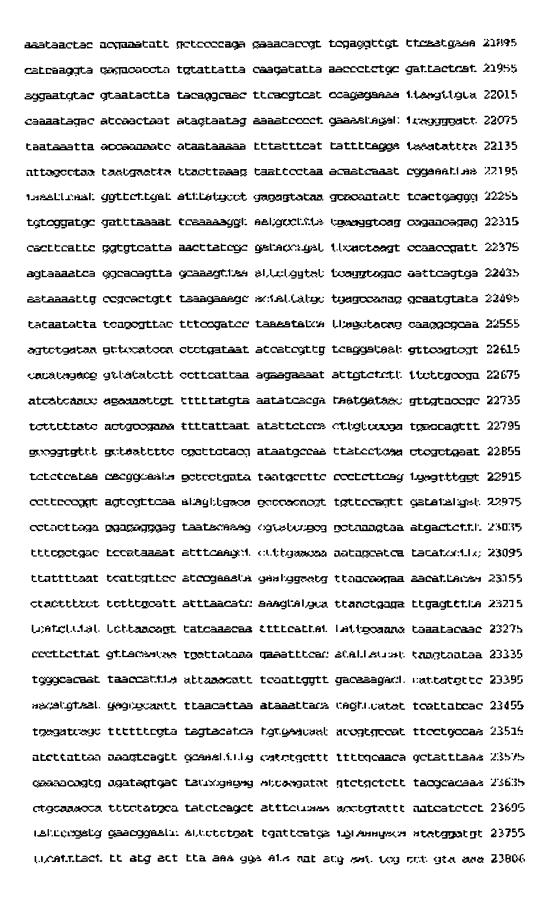
um Ser	gog Ala	eet Pro	tog Sær	eca Pro 695	aca Thr	aat ABD	gcc Ala	ecce Thr	ങ്ങൾ കടമ 700	gta Val	yja gæg	ata 11e	aec Asn	atr Ile 705	Arg Cgc	17288
									occ Pro							17336
									gca Ala							17384
tog Ser	990 Gly 740	toa Ser	to: Ser	ald: 11e	ardi Titur	get Ala 745	cca Pm	etg I <i>e</i> ar	agt Ser	ece Pro	gta Val. 750	(3) Y 9035	aat Asn	aaa ശ്രേങ	tet Ser	17432
									act Thu							17490
									tca Secr 780							17528
									agt Ser							17576
									atg Yet							17624
		Ile							aet Aen						gtic Val	17672
	Ason								cca Pro						tca Ser B50	17720
act Thin	tac Tyr	ata Ile	aaa Lys	tat Tyr 855	Thr	aag Lye	ysb	ooa Lye	tet Ser 360	aca. Thr	gta Val	tgg Trp	gtc Val	tet Ser B65	act Tu:	17768
				Glu	Ala	Gly	Gly	. CJII		Ser	Gly			دخعة	rat Ris	17816
			M et					Phe					Gju		cta Leu	17864
		Leu					Ser					Pro			(:Ca Leu	17912
	Asp					Glu					Ile				саt 813 930	17960
					Āla					رصا:					ctt Leu	18008

WO 99/42589

18055

aan aat gtg een eet tot aag aga tan egammaatta atattettta Nys Asn Val Lys Pro Tyr Lys Arg 955

tetaetttta atageeetet tgaacttaca eteaaggagg ggazzeeaaa taagaaacea 18115 teletastan esaceratga aagsatatti attienigge tigattaett itaarattea 18175 gtattagg() stiemmerm taicimmerm atimemetam caminetta ilimitatemet 18235 Leassietca aatcagaaat taatgaatit aagggiteit tatzittati teigagagea 18295 taggeseast accttacoga taggesetasa estaatteaa aateeagaaa tagtatatti 18355 testeaatat gogragusta gografitos tigogagica tiasactiat egogagacee 18415 octitized gatoczatet altaglasas traggosory licastazuje tasatititga 18475 tattosggga gataattosa tygoatsaaa ttattooatt gittitaaaaa agoactatta 18535. tgotgaacna aaggamasut egalialilati komicegogi geoliitotgg tiotaanata 18595 teatoggata cagcasgaga cagcattiga taagcaccat ciatocigat gatatoatea 18655 ttaluttoost aacattoogt ogtoocataa actottatat cocctttoot Lagongagag 18715 aatarrycat ottgoottat taaatoatoa attagaaaat tyttyattat araaatatog 18775 ogataatgat eacyttycau cyctotitti acyaccytay etaittiatt aacataitei 18835 ccactigige castascos titigicies titigatasti tataattiet acgaraatte 18895 caattottet caacttteag gateetttea taacaeggea geaactettg atatagtgee 18955 ttteeetett etgigagett ggittitteee ggiagteget eaaacaattg acaeeeraca 19015 omotyttoca gitagatatan gagootyota agiggagaag gggtamitaca aagegtatee 19075 geogotaatg bgaatgackii Mikoliiagol galkonakaa aatactttag ttgotttgaa 19135. cassatates tescatacee tettytiite atteesgass tagastutta accatagaan 19195 atgacascys tytticiant thycethott thecallegy acatgogtta atgymcatty 19255 autticacts catcastigt tastatibat Lieatectig coccatanit stamminam 19315 tataectteg tteattatti: citgatatte atvatogtea ettitocica ataectacae 19375 aagtagatal 148fillaket toragtaato tatogtitigg ogaoggaggi egaticitice 19435 attgggstat tomaccoatt ogoogootil ollaktaatt acagtgattt ttggcatttt 19495 gglittrator zacttaggit lataggtgat lillrowthta geaccoggig tizacttoan 19555 cotaaaggga tarstaceaa etteacettg taagaatatt etgtttggte tacetteaac 19615 gactiticama atggggtama taaccegggti aamatcamic gtatecamic catcamitte 19675. gotgatatit yteogggeig esteatigat aaatgegati aaateggtiy elgaalaceg 19735. satagratet tipactagat garggarate ggtataacte actgaracea aggrixggix; 19795. astritorac itacatogae egecacesti azzaggiagi itigerigaa agiaserggi 19855 titoggates getillisesi. ersejäugtaa kongulistaa getiggiaeet taasaggega 19915 catattopaa totaaaogat ahluaaggos ahottitigag atalacadag oggatacatg 19975 eggetgtgtg talltagglig cyanichtic iacagtamic cactgatict cliteggasy 20035 apagagegge teatttgogt (ageacages tgatasteaa atsacagata agazagataa 20095) quanticing etaittatca teglaselit tenteseno classeogit atotocatot 20155 gigteenatt compatette eigitimist attingement mamitageta egetgatage 20215 attactical alificatac atgaategaa aategactic tigagigging (talwaalit 20275 topogramy staticeacc racingeogn etilicitati agicaccyty auchtrycha 20035 tthtogtttc atomagotta ggottaaaaa taattttoos tttagotoot ggagttaacg 20395 bynyttyssa syyskyksti titastakii eskettytäs yastattety tikyyykyky 20455. ottowecyan itimaasama gyytääätää onggontasa atmaatoyia ilumahtios 20515. agatitigi aataticato tyyactatyo cattyataya tyogattaaa qoygilgoty 20575 aetacqqaat agcatottto accegniggo igacatoagi ataacicaco gaiacaaagy 20635 ecogyttaat tittooatita entegioeee elooatiaaa agytagitti getigaasai 20695 aaccogsttog tagatogeco thoachitos gacqaagcou attataggto ggoactitaa 20755 azgropaczt attęgastoc sęschatack chargecaste etttgatatg tattetgegg 20815 atacabgtgg Ulxggtstat htmggcgdta cocchictae cybgatocat tgattittett 20875. Raggaggga aageggetem titigggtem caeageetga tallaaaate setgmemma 20935 caaataagta tittittaaca titatosigg taagiittico toaattoois ongoattato 20995 operatorate tectyterag antagogtte attgattteg tercerarge raccargatay 21055 todosalcci: ettaccacag atasasasca cogctlabgo ogtgagtast agtgagtiga 21115 gogacaggga tacagcagtg catcoccatc aattagtooc titigaataaa gggaacagaa 21175 tttgsaacut cogtcatecc gtccstatta cygascttag attatgatta ttaaatcacc 21235 accadalggc asgessett ttcattttt settlacea gastgaattt qtaaqasagt 21295 gttakaaant taatagaaat taatiteetg itaalotant gaaggatgaa attataaaaa 21355 teacccatti cicagggaca acestocaca atatétages coactogico icacitaati 21615 topbytosey agtagaasta teetgatgae teagloopaty acataczgoa atqteattgq 21075 tattgegect accessigit taateesttt otttbytott taatggegeg etecsegige 21535 ttoactatti dagoactate gataaataag attoopaaat agogocatat ettecaemae 21595 toataattul aigtalaaco attgyttäää tapgalootg tytääcägga tiatgaaakog 21655. ttatttatet casaterate sattattita tatatagitt caragicars chris(salci: 2)7],5 pytacettea taaceaacty (cotocotyc petacettet gataacaaca getacactaa 21775 ctataccogo gmctataatt atgaccytyb gaaattoog cytagttoac cygccacyca 21835.



Mat Ile 1mu bys Gly Il Asn Met Asn Ser Pro Val Typ. 960 965

gay eta out gat gta tta asa atu day lgt ggt tit dag tgt dig ad Glu Ile Pro Asp Val Leu Lys Ile Glu Cys Gly Phe Glu Cys Leu Th 970 975 980	a 23854 r
gat att age cae age tot tit Aan gaa lut cae cag caa gia toe ga Aap Nie Sar Bis Sor Ser Pha Ash Glu Pha His Glu Glu Val Sor Gi 985 990 995 1000	ī.
cae ete tee tog tee goa goa eae gae tta tat eet gat gen ean eag His Lau Ser Tro Ser Glu Als His Asp Leu Tyr His Asp Ala Glu Glu 1005 1010 1015	g 23950 o
god cas eag got sat ogg otg tat gas gog ogt ett ott ann ogd ac Ala Cln lys Asp Asu Arg Leu Tyr Clu Ala Arg Ile Leu Lys Arg Thu 1020 1025 1030	⊋ 23998 r
aat oot ean the can aat got gta cat oft god atc gte gog oot aar Am Pro Gin Leu Gin Am Ale Vel His Leu Ala fie Val Ale Pro Ag 1035 1040 1045	240 4 6
get gas etg ata gge tat aac caa tit age gge agg gem agt caa Ala Glu leu Ilo Gly Tyr Asn Asn Glo She Ser Gly Ary Ala Ser Glo 1050 1055 1060	a 24094 a
tat gto gon con ggt acc gtt tro toc atg tto toc occ gcc gcc to Tyr Vel Ale Pro Gly Thr Vel Ser Ser Net Sie Ser Pro Alo Ala Tyr 1065 1070 1075 1080	r
ttg act 9ag ott tat ogt gea gos ogc eat tta cac god agd get to ieu Thr Glu Leu Tyr Arg Clu Ala Arg Asm Leu His Ala Ser Asp Ser 1085 1090 1095	; 24190 r
yd.E tat ege etg gat act ege ege eea gat etc aaa tea atg geg etc Val Tyr Arg Leu Amp Thr Arg Arg Pro Amp Leu tym Ser Met Ala Leg 1100 1105 1110	2423R
agt cas cas ast alg gat acg gas ett tec act mic tet tta tec ast Ser Cln Cln Asn Met Asp Thr Glu Leu Ser Thr Leu Ser Leu Ser Ag 1115 1120 1125	t 24286
gag ots the the gam ago att ame act gag hold may one get man that Glu Leu Leu Clu Ser Ile Lys The Glu Ser Lys Leu Asp Aum Tyn 1130 1135 1140	t 24334
Set cas gtg atg gas atg etc tec get ttc egt ecu Leu gge geg acq Thr Cln Val Met Glu Met Leu Ser Ala Fho Arg iro Ser Gly Ala Thu 1145 1150 1155 1166	; ;
oct tat dad gat got tad god dat gtg ogt aaa gitt abo oog ota daa Pro Tyr Bia Asp Ala Tyr Glu Asm Vol Arg Lys Vol Ile Glu Lou Glu 1165 1170 1175	2 44 30
got oct gag ott gag cas tla aal got tox oca god att god geg ott Acp Pro Gly Leu Glu Gli Leu Asm Ala Sen Pro Ala Ile Ala Gly Lei 1180 1185 1190	3 2447B
atg cat can get tee eta tim ggt att ace get tea ate tea eet geg Met Nis Gin Ala Ser Iau leu Gly Ile Asn Ala Ser Ile Ser Pro Gin 1195 1200 1205	3 24526 I
ttg til AAC abt otg ocg gag gag att act gan ggt aat get gag gas las Phe ABN lle Leu Thr Glu Glu lle Thr Glu Gly Asn Ala Glu Glu	a 24574

1210	3215		1220	
ctt tat ang AAA Leu Tyr Lys Lys 1225	aat tiit ggl: Asn Yhe Gly . 1230	eat ato gwa ong Aso lle Glu Pro 1235	get toe etg get atg Ala Ser Leu Ala Med 1240	
Fro Glu Tyr Leu	aga egi, tai; Arg Arg Tyr 1245	tad ast tha agu Tyr Aøn Leu Ser 1250	. 981 988 ඉහස ete age Amp Glu Glu Izas Se 1255	24670
cag ttt att ggt Gln The De Gly 1260	aaa god ago Tys Ata Sor .	aat tto ggo caa Asn Pho Gly Gln 1265	. C&A gas tat agt am Gln Glu Tyr Ser Asi 1270	24718
aac caa ete att Asm Glm Lou TJe 1275	That Prox Illo '	gto aac ago aat Val Asn Ser Asn 280	gat ggc acs gtc ear Asp Gly Thr Val Lyr 1265	2 476 6
gLa fat oga att Val Tyr Arg Ile 1290	acc cgc gas Thr Arg Clu 1 1295	Tyr Thr Thr Asn	gnc aat caa gta gad Ala Asm Glm Vel Asg 1300	24814 >
gty gag cty tit Vel Giu Leu Phe 1305	ecc tac ggt : Pro Tyr Gly : 1310	gqa gaa aat tat Gly Glu Aan Tyr 1315	cag the agu tee cas Gin Leu Amn Tyr Lyd 1320	3
Phe Lye Amp Ser	ogt oag gat : Ang Glin Asp ' 1325	gto too tat tta Val Ser Tyr Ieu 1330	ter ate esa tia esi Ser Ilo Lys Leu As 1335	; 24910 1
gac aan aga gaa Asp Lys Arg Glu 1340	Leu Ile Arg	att gome ogga grog Ile Glu Gly Ale 1345	ert eag gir aac ato Pro Glo Val Aeo Ile 1350	24958 e
gaa tat tea gaa Clu Tyr Ser Glo 1355	His Ile Tm	ita agi ara act Leu Ser Thr Thr 360	gut atc agt caa ce Asp Ile Ses Clu Pro 1365	. 25 00 6
ttt gaa ate gge Phe Glu Ile Gly 1370	eta aca ega Leu Thr Arg 1375	Val Tyr Pxo Scr	agt tel bog goa tal 'Ser Ser Trp Ala Ty: 1380	: 25 05 4
gca gro gca waa Ala Ala Ala Lye 1385	ttt acc att Phe Thy Ile 1390	gag gaa tat aac Glo Glo Tyr Asn 1395	caa han bot tto otg Gin Tyr Ser Phe Lea 1400	i
Leo Lys Leo Aen	Ama got att lys Ala Ile : 1405	egt eta tet egt Arg Leo Ser Arg 1410	geg aca gea tta te: Ala Thr Glu Leu Se: 1415	a 25150
cxx: ecx: att ctg Pro Thr lle Leu 1420	Glu Ser Ilo	gig ogt agi git Val Arg Ser Val 1425	aat cag cea cly gar Aso Glo Glo Leu Mq 1430	: 25198 :
ate aac gea gan He Asn Ala Giu 1435	Val Leu Gly	asa gil tuu eug Lys Val Phe Lea 440	not man tat tat ato The Bys Tyr Tyr Med 1445	3 252 46 :
csa cgt tat gct Gin Arg Tyr Ala 1450	att ast get : The Asm Ala : 1455	Glu Thr Ala Leo	nta ota topo aat gos Ile Leu Cyo: Ayo Alo 1460	35294 U
ott att toa toa Loo The Ser Ghn 1465	ngt toa tat (Arq Ser Tyr) 1470	gat aat daa oot Asp Asn Gin Pro 1475	Agu twa ttt grit cry Ser Glin Phe Asp Ari 1480	,

ntg tit aat acg oom tim oig een gge omm lan U.A. bel mee gga gen. Leu Phe Per Thr Pro Leu Leu Aen Gly Gln Tyr Phe Ser Thr Gly Aep 1485 1490 1495	25390
gas gag att mat the ear one mut agt act uge out too ear eas too the Clu The Asp Lea Ase Pro Gly Ser The Gly Asp Trp Ary bys Ser 1500 1505 1510	25438
yty off am oyt gom tit mat ato gat gat att too ofe tae typ ofg Val lam Lye Arg Ala Phe Am Ile Amp Amp Ile Ser Len Tym Arg Len 1515 1520 1525	25486
ott ase 311. soc 25c cat 25t. 25t. cas get gga 25g Alt. 24a 25t. 25t. Leu Lys 11e The Ase His Ase Ase Gle Asp Gly Lys file Lys Ase Ase 1530 1530 1536	25534
lie eet ast oft fot gat fits fat ein gog eae fits mig gos ges eit Leu Ash Ash Lou Str Asp Leu Tyr lie Gly Lye Leu Leu Ala Glu lie 1545 1550 1560	2558 2
cat cas tta acc att gat gas ttg gat tta ttg ctg git god gig ggt His Gin len Thr lle Asp Glu Len Asp Len Len Vol Ala Val Gly 1565 1570 1575	25630
gaa 990 995 act. ast tha ton got ato agt gut ann cas otg gog gos Glu Gly Glu Thr Asn Leu Ser Als Ile Ser Asp Lys Gln Leu Ala Ala 1580 1590	2S67H
ctg atc agu aks étc aat acc att acc gte tog eta eag aca eag ang Leu lle Arg Lys Leu Asn Thr lle Thr Val Tro Leu Cln Thr Gln Lys 1595 1600 1605	25726
tag sg); goy tir cas tha tit got sig sot toe see age tal asc ass Top Ser Ala Uns Gin Leu Phe Val Met Thr Ser Thr Ser Tyr New Yor 1610 1615 1620	25774
acg ctg acg cct gaa att aag aat ctg ctg gal acc glo; tec cec ggt. The Leu The Pro Glu Lie Lym Asn Lou Leu Asp The Val Tyr His Gly 1625 1630 1635 1640	25822
tta caa ogc tit gat aas gad aæg gds ast tie olg dat git atg gdg Lou Gin Gly Phe Asp bys Asp bys Ala Aæn Leu Leu His Val Met Ala 1645 1650 1655	25870
eec tat att gog goe ace tta caa tta tea teg gas mal gto gee oat Pro Tyr Ile Ala Ala Thr Leo Glo Leo Ser Ser Glo Aso Val Ala His 1660 1665 1670	25918
tot gly ole off. Lyg you gad and the and ecc ggd gad ggd gros alg Ser Val Leu Trp Ala Aep Dys Len Lys Pro Cly Asp Gly Ala Met 1675 1680 1685	25 9 66
aco see son and the top year top they ask are can less eas son occussor. The Ala Glu Lys Pic Trp Asp Trp Leu Ash Thr Glu Tyr Thr Pro Asp 1690 1695 1700	26014
toa tog gaa gta tha gos ace meg gea cest ett gtt can tat tgt can Sor Sor Chu Val Lou Ala Thr Gin Glu His Ile Val Chn Tyr Cys Gln 1705 1710 1735 1720	26062
QCG ttg QCG coa tta gan atq qtt tac cat tcc acc gg), aus sai gas Ala Leu Ala Glu Leu Glu Mei: Val Tyr His Scr Thr Gly ile Ash Glu 1725 1730 1735	26130

WO 99/42589

asc gor the ego ong thi grig aca saa ona gag ang thi ggo tog now Asm Alm Phs Ang less Phe Val Thr Lys Pho Glo Net The Gly Ser Ser 1740 1745 1750	2615B
act, gag goa gta oot gog oat gat goa ott toa otg etw atg otg acg Thr Glu Ala Val Pro Ala His Asp Ala Leu Ser Leu Ile Met Leu Thr 1755 1760 1765	26206
ogu tit goa gat tog git aat gog tia ggo gas eaa god tot too gia Ang Phe Ala Asp Tro Vol Asm Ala Leu Cly Clu Lys Ala Ser Ser Vol 1770 1780	26254
cta gog goa itt gan got aac agt tia acg gon gan can tig got gat Lou Ala Ala Pre Glu Ala Amn Ser Len Thr Ala Clu Gln Len Ala Amp 1785 1790 1795 1800	26302
god atg aat ett got got aat tig die itig daa god agt eel daa god Ale Met Asm (an Asp Ala Asm (an Leu Gin Ala Ser Thr Gin Ale 1805 1810 1815	26350
cae aac cat cae cat off occ con gtg acg cae aae eaf got fite too Gln Aen His Gln Ris Leu Pro Pro Val Thr Gln Lys Asn Ala Phe Ser 1820 1825 1830	26398
tgt (gg sus tot ato gas act atr ctg caa tgg gtt aat gtt gca caa Cym Trp Thr Sar Ile Asp Thr The Leu Gln Trp Val Asn Val Ala Gln 1835 1840 1895	26446
cas tig sat gir ger ona cag ggs gil tox get tig gir ggg cig gat Gln leu Asn Vol Ala Pro Gln Gly Val Ser Ala Leu Val Gly leu Asp 1850 1855 1860	26494
tat att can tto aat can aan atc occ acc tal goo can tog gan agt Tyr Ile Glo Leu Aso Glo Lys Ile Pro Thr Tyr Ala Glo Trp Glu Sar 1865 1870 1875 1880	26542
get ggg gas ats tig ark, ged ggg tig ast tes cas cag get gat ats Ale Cly Clu Ile Leu Thr Ale Gly Leu Asm Ser Glo Glo Ale Asp Ile 1885 1890 1895	26590
tta cac get til tig gad gam kui ogn mgi ged gem ita agd mud tmd Leu His Alm Pom Leu Amp Glu Ser Ang Ser Alm Alm Lou Ser Thr Tyr 1900 1905 1910	26638
tat ato ogt caa gto gov eng oce gog goe god ata aan ego ogt gat Tyr lle Arg Cln Val Ala Dys Pro Ala Ala Ala lle Lys Ser Arg Asp 1915 1920 1925	26686
ged tig isd cas isn lis dia att gat dat dag git idd god atd Amp led Tyr Gin Tyr Led Led Ile Amp Amn Gin Val Ser Ala Ala Ile 1930 1935 1940	26734
asa act ace ogg att god god god att god att cas cly tec gto Lys Thr Thr Ary The Ala Glu Ala He Ala Sor The Gln Leu Tyr Val 1945 1950 1955 1960	26782
asn. Opc and Cly gam and glas gas gas ast god cal the gag git atc Asn and The Leu Glo Asn Val Glu Glu Asn Als His Ser Gly Val Ile 1965 1970 1975	26830
age ogt omg tte tit ate gae tgy gan aam hwi ame man oge tae age Ser Arg Gln Phe Phe Ile Amp Top Amp Lym Tyr Amn Lym Arg Tyr Ser 1980 1985 1990	2687R
wax long god got got tot oza lla gott tad tad eeg ges aan lat. att	26926

WO 99/42589

That Tags Als Gly Val Sem Gla Lew Val Tyr Tyr Pro Glu Asm Tyr Ile 1995 2000 2005	
gat occ ecc atg ogt atc ggs cam acc sam etg etg gen gom tha thy Mep Pro Thr Met Arg Ile Gly Glm Thr Lys Met Met Asp Alm Leu Leu 2010 2015 2020	26974
can too gite ago cam ago cam bha amit goo gat ach gite gaa gan gon Ghn Ser Val Ser Gin Ser Gin Leu Amn Ala Amp Thr Val Glu Amp Ala 2025 2030 2035 2040	27022
ttt abg tot tat ofg aca bog bit gag oan gig geb aat obb ana gib Pha Met Ser Tyr Leu Thr Ser Phe Glu Glu Val Ala Asu Leu lys Val 2045 - 2050 - 2055	27070
att agn gog fæt dad gaf, sat att aæd aæd gæt daa ggg dig add tal Ile Sen Ala Tyn Ris Aæp Asm Ile Aem Aem Aep Glm Cly Lou Thn Tyn 2060 2065 2070	27118
(ii), and ggm ofc agt gas act gat acc ggt gas tac tat tgg ogd agt The lie Gly Lew Ser Glu Thr Asp Thr Gly Glu Tyr Tyr Trp Arg Ser 2075 2080 2085	27166
yte gat cac agt amma tte age gme ggt mamma tte gee get amm gee tag Val Amp Ris Ser Lys Phu Ser Amp Cly hym Phu Alm Alm Amm Alm Trp 2090 2095 2100	27214
agt gam tog two mem att gat tigt och att amt och the oga mig met Ser Glu Trp Him Lym Ile Amp Cym Pro Ile Amn Pro Tyr Arg Ser Thr 2105 2110 2115 2120	27 262
ato ogt och ýdg álg bao ama toc ogc tig tai eig etc igg itg gam Ile Arg Pro Val Mei Tyr Lym Ser Arg Leu Tyr Leu Leu Trp Leu Glu 2125 2130 2135	27310
caa sag gag atc act aas can acn gyn aat agc ama get gge tat ena Gin Lys Clu Ile Thr Lys Gin Thr Gly Asn Ser Lys Asp Gly Tyr Gin 2140 2145 2150	27358
noe gag aca gat tat ogt tat gag eta aaa tig geg eat ate egi tat Thr Glu Thr Amp Tyr Arg Tyr Glu Leu Lys Leu Aia Bis Lhe Arg Tyr 2155 2160 2165	27406
gae ggt aux bgg dat acg cea atc act tit gat gic aat gaa aaa ata Asp Gly Thr Trp Amn Thr Pro Ile Thr Phe Asp Val Ash Glu Lys Ile 2170 2175 2180	27454
too aay ista gaa oti goo aan ant ann gog oot gog oto tat tigt got Ser Lym Leo Glo Leo Ala Lyn Asn Lyn Ala Pro Gly Leo Tyr Cys Ala 2185 - 2190 - 2195 - 2200	27502
ggi, tai: cea ggt gea get mog tig cig git aig tit tai mec cea cea Gly Tyr Gln Gly Glu Asp Thr Leu Leu Val Met Phe Tyr New Gln Gln 2215	27550
gat now ofto get agt tat ama acc get tem atg cam ggg eth tat atg Agy Thr Leu Asp Ser Tyr Lys Thr Ala Ser Met Gln Gly Leu Tyr 1le 2220 2225 2230	2759H
ttt god gut atg gam tat mam gmt atg mod gat ggm dam tad mam fæll. Phe Alm Amp Met Glu Tyr Tyr: Amp Mot Thr. Amp Gly Oln Tyr Tym Son 2235 2240 2245	27646
IAC cgg gad aac ago tat aww cam tto gal wot wat agt gto aga aga Tyr Arg Axp Axp Ser Tyn bys Gln Phe Asp Thr Axo Ser Val Arg Arg	27694

2250	2255	2260	
gtg WAL BAC CGC Val Asn Asn Arg 2265	tat goa gog Tyn Ala Glu 2270	gat tat gam att ooc toe tog gin aat Asp Tyr Glu ile Pro Ser Ser Val Asn 2275 2280	27742
Ser Arg Lys Gly	tat gat tgg Tyr Asp Trp 1285	gga get tot tot etc agt atg gts tat Gly Asp Tyr Tyr Leu Ser Met Vel Tyr 2290 2295	27790
aac gga gat att Agu Gly As p Ile 2300	Fro I'm 1le	agt tat ama goo ach ten agt gat tta Ser Tyr bys Ala Thr Ser Ser Asp Leu 2305 2310	27838
ass also tat atc Lys lie Tyr lie 2315	Ser Pro Igo	tta aga att att cat aat ggs tal дэн Teu Arg Ile Ile His Asa Gly Tyr Glu 2325	2788G
ggg cay caa ogc Gly Gln Gln Arg 2330	aat eaa tge Asn Olo Cys 2335	aat ota atg aat aas tat ggs: waa ota Asn Lon Mot Asn Lya Tyr Gly lys Leu 2340	<i>2</i> 7934
ggt gat aaa ttt Gly Asp Ly: Phe 23 4 5	att gtt tat Ile Vai Tyr 2350	act age tig gga git aat oos aak aat The See Leu Gly Val Ash Pro Ash Ash 2355 2360	27982
Sor Ser Man Lys	ctg atg ttt Leu Met Phe 2365	two occ gtt tat cam tat amc ggm est. Tyr Pro Val Tyr Gin Tyr Asn Gly Amn 2370 2375	28030
gtc agt ggg ott Val Ser Gly Leu 2380	Ser Gin Gly	aga tta eta tte ese egt gas acc aat Arg Leu Leu Phe His Arg Asp Thr Asn 2385 2390	20078
tat toa uni aas Tyr Ser Ser Lys 2395	Val Ciu Ala	tộy átt cơi gọa gọa gọa cại tọi cia Trp Ile Pro Gly Ala Gly Ary Scr Low 2405	28126
acu Ast cog ast Thr Asn Pro Asn 2410	get ged att Ala Ala Ilo 2415	ggt gat gat hal got ace gno tog tha Gly Asp Asp Tyr Ala Thr Asp Ser Les 2420	28174
aac Aea oog aat Asn Lys Pro Asn 2425	gat ott aag Aap Leu Lys 2430	cea tan gin tat atg act gas agt aaa Gln Tyr Val Tyr Met Thr Asp Sen Lys 2435 2440	28222
ggt act gct acc Gly Tou Ala The	get gto toa Asp Val Sor 2445	ggs cca gta gs(atc ast act gca att Cly Pro Val Amp The Amm Thr Als Ile 2450 2455	28270
tee eeg gym waa Ser Pro Ala lys 2460	:Val G)n Ve(aca gta ama gee ggt age ama gas mas Thr Val Lys Ala Gly Ser Lys Glu Gln 2465 2470	2831B
arus litit acc geg Thir Phe Thir Ala 2475	. Aleop Liya Aleo	gic hoc all rag coa toe cot ago tit Val Ser Ile Glm Pro Ser Pro Ser Pho 2480 2485	28366
gat qua atg aat Ag) Giu Met As: 2490	tat coa ttt Tyr Gln Phe 2495	. aat get ete gaa ata gat gge era agl. : Asm Ala Leu Glu Jie Asp Gly Ser Ser 2500	28414
otg am. tit act Leu Amn Pho Thu 2505	. asc sat toa · Asn Asn Scr 2510	god agl act gat att act tit acc goa Alm Ser fle Amp lle Thu Phe Thu Ala 2515 2520	28462

WO 99/42589 PCT/EP99/01015

ttt gom gang get gge ogt amm otg tigt tat gam agt tto agt att oot Præ Ala Glu Asp Gly Arg bys kom Gly Tyr Glu Ser Phe Ser Ile Pro 2535 2535	28510
att ace oge meg gbg egt ext get aaf tee etg ace etg ege cas mat lle Thr Arg Lys Val Ser Thr Amp Amp Ser Leu Thr Leu Arg His Asn 2540 2545 2550	28558
gsa aat ggt gog cee left aky cea tyg ggs gto tat ogd att ogt ott Ghn Asa Gly Ala Glin Tyr Met Glin Trp Gly Val Tyr Arg Ile Arg Lox 2555 2560 2565	28606
ant ext. the fit get eye can the git geg egg gee act acc ggt aft Amo für Leu Pic Alm Ang Glu Leu Val Ala Ang Ala für für Gly Ile 2570 2580	2865/1
gat why aft ctg agt atg gas act cag sat att cag gas com cag tta Amp Thr 11m Low Ser Mot Clu Thr Gin Amn Ile Gin Gin Pro Gin Lew 2585 2590 2595 2600	28702
ggr ama ggt tte toe get mog tte gtg mtn eet eeg tat mae tem Gly Ing Gly Phe Tyr Alm Thr Phe Val Ile Pro Pro Tyr Amn Pro Ser 2605 2610 2615	28750
act cut ggt gau gas cgt tgg Ltt sag cth tal atc sas cat glu gli: Thr Ris Gly Amp Glu Arg Trp Phe hys Leu Tyr Tle hys His Val Val 2620 2625 2630	28798
gath ast and the cat att ath the test the ggt head of the assignment as as as and asp Asm Asm Sec His III III Tyr Ser Gly Glin Leu Lys Asp Thr Asm 2635 2646 2645	28845
ata ago ace aog tha itt ato cot off gat gat git coa thg aac coa Ile Ser Tur Tur Leu Phe Lle Pro Leu Aep Aep Val Pro Leu Aen Q(π 2650 2655 2660	28894
get tod agu gru dag git tod atg acc tid mag man tom uum tom gat Aup Tyr Ser Ala Lys Val Tyr Met Thr Phe Lye Lye Ser Pro Ser Asp 2665 2670 2675 2680	28942
ggt acc togg figg ggc oct oac tit gut aga gat gat aan gga ata gta Gly Thr Trp Trp Gly Pro His Phe Val Arg Amp Amp Lym Gly Ile Val 2685 2690 2695	28990
ace ata eet CCL ees toe att tig ace dec tit gag age gio eet gio Thr lle Amn Pro Lys Ser lle Leu Thr His Phe Glu Ser Val Ash Val 2700 2705 2710	29038
etg pat aat Wil agt age gom een atg got tte ege gge get ome oge Leu Asn Asn Ile Ser Ser Glu Pro Met Amp Phe Ser Gly Ala Amn Ser 27)5 2720 2725	2908 6
ete tat itt igg ges etg ite tae tai ace etg afg etg git gee eaa Leu Tyr Phe Trp Glu Leu Phe Tyr Tyr Thr Pro Mei Leu Val Als Cin 2730 2735 2740	29134
egt tig tig eat eag caa aac tit gat gas geg aac ege igg eig aas Arg Leo Leo His Gio Clin Asn Phe Asp Glu Ale Asn Arg Trp Leo Lys 2745 2750 2755 2760	29182
tat sto tog ogo oco too gog tat att ott oec ogo cao att cas aat Tyr Val Trp Ser Pro Ser Gly Tyr Ile Val Sis Gly Glm Ile Glm Asm 2765 2770 2775	29230

tst caa tgg aac gtc ogc oog tta ttg gaa got occ agt tgg aac agt Tyr Glo Trp Aso Val Arg Pro Leu Leu Glu Asp Thr Ser Trp Aso Ser 2780 2785 2790	29278
gat cet tig gat tee git get eet gae geg gin geg eag eac gat eeg Asp Pro Leu Asp Ser Val Asp Pro Asp Ala Val Ala Glim His Asp Pro 2795 2800 2805	2 9326
atg one tat aan got toa ace tot atg oge ace off gat ofg tig atc Met Nie Tyr Lys Val Ser Thr Five Wet Arg Thr Leu Arg Len Len (1): 2810 2820	29374
gog cyc gg: gar: cat gct tac cyr: caa the gay cyr gat acg ctt aac Ala Arg Gly Asp His Ala Tyr: Arg Gln Leu Glu Arg Asp Thr Leu Asp 2835 2830 2835	29422
gae goy aag atg tog tat atg caa goy otg oat otg tte ggo gat eans Glu Ala Lys Met Tro Tyr Met Gln Ala Leu His Leu Leu Gly Asp Lys 2845 2850 2855	29470
oct hat only only difference and logy wat day one ogs only gad ass Pro Tyr Leu Pro Leu Ser Thr Thr Trp Amn Amp Pro Arg Leu Amp Jys 2860 2865 2870	29518
god gry gat att art are daa agt get dat ter agn kom atm gte get Als Ala Asp Ile Thr Thr Gln Ser Ala His Ser Ser Ser Ile Val Ala 2875 2880 2885	29 566
ttg tag cag agt dea teg gog oft tha toa thg ogc age gen agt dee Leu Ang Gin Ser Thr Pro Ala Lou Lau Ser Leu Ang Ser Ala Agn Thr 2890 2895 2900	29614
ots acc gut oto the otg ocg cam ato eat gea gty atg ang ant two Lew Thr Aep Lew Phe Lew Pro Glo Ilo Aso Glu Vol Mat Met Aeo Tyr 2905 2910 2915 2920	29662
tgg caa acs tta gef cag aga gta tac aac etg ege cac aac etc tet Trp Gln Thr Leu Ala Gln Arg Val Tyr Aen Leu Arg His Aen Leu Ser 2935 2935	29710
ate gae ggt tag tog tta tat etg eea ate tat ger; aca cog gog gae Ile Aep Gly Gln Pro Teu Tyr Lau Pro Ile Tyr Ala Thr Pro Ala Aep 2945 2950	2 9758
oog aan gog tta ete age goe get git goe eet tel: wax 9gt 9ga gge Pro bys Ala Leu Leu Ser Ala Ala Val Ala Thr Ser Gln Gly Gly 2955 2960 2985	29806
awy olg odg gwy ton tit atg too etg igg ogt ito ong omn mly otg Lys Leu Pro Glu Ser Phe Wel. Ser Lou Trp Arg Phe Pro His Met Leu 2970 2980	29854
gwa aat get ege age atg git age eag ein ace cam the gge tee meg Glu Asu Ala Ang Ser Wet Val Ser Glu Leu Thr Glu Phe Gly Ser Thr 2985 2990 2995 3000	29902
tta caa aat att atd gee dyb dag ged god ged gog etd det geg tta Lou Glo Aso He He Glu Arg Glo Asp Ala Glu Ala Leu Arm Ala 1km 3005 3010 3015	29950
tta caa eat cag god goa gag dig aba lig ant wad dig agt att caa Ieu Gio Asn Gin Ala Ala Ghu Leu Fle Leu Thr Asn Leu Ser Ile Gin 3020 3025 3030	29998
gae ann ach all gas gas etg gat ged gag ann ach gtg etg gas man	30046

Asp Lys Thr fle Glu Glu Leu Asp Ala Glu Lys Thr Val Leu Glu Lys 3035 3040 3045	
too awa gug gga gna cas tog ogo tit gst ago tat ego eka obg cat Ser Dys Ala Gly Ala Gln Ser Arg Phe Asp Ser Tyr Ser Lye Leu His 3050 3055 3060	30094
gat gam sen ato eac god ygh gwa sen nas yoh atg acg cta cga yog Asp Ghu Asn lle Asn Ala Ghy Ghu Asn Ghn Ala Met Thr Leo Atg Ala 3065 3070 3080	30142
time gos god gog ott see seg god got dag god Min ogt otg god god Ser Ala Ala Gly Leu Thr Thr Ala Val Glin Ala Ser Ang Leu Ala Gly 3085 3090 3095	30190
qua que got got cue cue gre cet amo ato the gre the god got got got Ala Ala Ala Asp Leu Val Pro Aso I.l.o Pho Cly The Ala Cly Gly Gly 3100 3105 3110	30 238
age egt tgg ggg get ate get gag gog ace gge tat gta atg gaa tit Ser Arg Trp Gly Ala Ile Ala Glu Ala Thr Gly Tyr Val Met Glu Fie 3115 3120 3125	3 02 B6
tox: gc), eat gtt atg est eox: gae gog got aoa att agc caa tot gae Ser Ala Asn Val Mat Asm Thr Glu Ala Asp Lys Ile Ser Glu Ser Glu 3130 3135 3140	30334
ace tae egt ogt oge egt eag gag tag gaa alt oag ogt aat aat goe Thr Tyr Arg Arg Arg Gin Glu Trp Glu Ile Gin Arg Asm Asm Ala 3145 - 3150 - 3155 - 3160	30382
gea gog geg otg aaa daa dio gal goo dax ott ama tog otg gom gla Glu Ala Clu Lou Lys Gln Leu Asp Ala Gln Leu Lys Ser Lou Ala Val 3165 3170 3175	30430
ego ogt gan god god gta tig das aas acc agd etg ann acc das ang Arg Glu Ala Ale Wal Jew Oln Lys Thr Ser Leu Lys Thr Glo Cln 3180 3185 3190	3 047 R
gang tang att (18% gate case ting gate til eing case ogt) asg tile aget aat Glu Glu Thr Glu Ala Glu Len Ala Phe Len Glu Arg lys Phe Ser Ago 3195 3200 3205	30526
cas gog the fac sac ton cts ent one eag one cty you gow att tac the Gin Ala Leu Tyr Asm Trp Leu Ang Gly Arg Leu Ala Ala Ile Tyr Phe 3210 3215 3220	30574
cam the two yaw the get ate des est tet the alg sea gas cas set Gin Phe Tyr Asp Leu Ala Ile Ala Arg Cyo Leu Met Ala Gin Gin Ala 3225 3230 3235 3240	30622
tad dyf tgy gaa aff. Ayn gat gad tot got dge tit all ama dog gge Tyr Arg Trp Glu ile Ser Asp Asp Ser Ala Arg line ile lys Pro Gly 3245 3250 3255	30670
gee tag can aga acc tat ges got atg ale ges ant ges acc tig sig Als Trp Gla Gly Thr Tyr Als Gly Leo Leo Als Gly Glu Thr Leo Met 3260 3265 3270	3ท77 ผู
cts agt tig goa cas alig gwa gae goe tet tia aga egn gwn aww ege Leu Ser Leo Alw Cln Met Cln Asp Ala His Leo Ang Ang Asp Lys Arg 3275 3280 3285	30766
goa tha gag gim gas ogn aca gia iog oig goo gas att tai goi gg. Als leu Glu Vol Clu Arg Thr Val Ser Leu Als Clu lle Tyr Ala Gly	30814

3290	3295	330u	
tta eeg eaa gat	ana ggo coa the t	too otg aog caa gaa	ato gag awg 30862
Lee Pro Ghi Asp	Sys Gly Pro Phe S	Ser Leu Thr Gln Glu	lle Glo tys
3305	3310	3315	3320
Leu Val Aan Ala	ggt toa gg: agn g	30: ggC agC ggt aat	eat aut tig 30910
	Gly Ser Gly Ser !	Ala Gly Ser Gly Asn	Asn Asn Sen.
	325	3338	3335
gos ttt gge ger	Gly ባቸው እ፷፬ ገኘው ፲	asa set tet tug dag	goaltoc att 30958
Ala Phr Cly Ala		Lys Thr Ser Leu Glin	Ala Ser Ile
3340		345	3350
tea tto get gat Ser Leu Ala Asp 3355	tta maa att tgt (Leu Lyn Jin Arg (3360	gag gat tac ccy gaa Mu Asp Tyr Pro Glu 3365	tet sil gge 31006 Ser Ile Gly
ana ato oge ogu Lys Ile Arg Arg 3370	ate bas cag ate a The Lys Oln The S 3375	age git ace etg mng Sur Val Thr Leu Pro 3380	90g ctaitty 31054 Alaleuleo
gga oct tan cag	gat gtg cog gca a	ata tia tet tac gge	gal awa goo 31102
Gly Pro Tyr Gln	Asp Val Gln Ala :	Tia Leu Ser Tyr Gly	Asp Lys Ala
3385	3390	3395	3400
Gly Leu Ala Asn	ggo tyb goa gog o	cty goo gtt tee cac	ggt acy ast 31150
	Gly tys Ala Ala I	Leu Ala Val Ser His	Cly Thr Aso
	405	3410	3415
gac agc ggt caa	Phis Clin Leu Asp 1	tte aac gat gge aaa	tte etg eng 31198
Aფა Ser Gly Gln		Phe Aen Aep Gly Lys	Phe Leu Pro
3420		425	3430
lith ges ygh ain Yne Glu Gly 11a 3435	ger att gat caa (Ala Ile Amp Glu (3440	ggt ang dis ecs ong Gly The Leu The Leu 3445	agt tit cet 31246 Ser Phe Pro
aat goa toa aog Asn Ale Ser Thu 3450	eca guu aan ggu : Pro Ala Lye Gly 1 3455	Dab caa god act atg Lys Gin Ala Thr Met 3460	CLA AMA ACC 31294 Leg Dys Thur
etg aac gat ate	ett Urg cat att (ogo tao acc att aag	tee 31336
teu Asm Asp Ile	lle Leu Nis Ile)	Ary Tyr Thr Ile Lys	
3465	3470	3475	
costocoso acaga	actau pacap geoco	gaateggggt etggtaa	ggw gittet aig 31395 Meg.
cag ant ton cag	aca tto ego alg m	His: gag ctg ten ita	oot aag gyr: 31443
Gln Aan Ser Gln	Thur Pha Seu Met '	Thr Glu Leu Ser Leu	Pro Bys Gly
3480	3485	3490	3495
CIA CIA VIO 116	acc ggt Alg ggt :	gaa goa tta acg eeg	god gyg nog - 31491
	Thr Gly Met Gly (Glu Ala Fau Thu Pro	Ala Gly Pro
	500	350S	3510
eat not ato dea	Ala Leo Ser Leo 1	oca tig occ att to:	gxx gga cgt 31539
Asp Gly Met Ala		Pro fæn Pro Ile Ser	Ala Cly Arg
3515		520	3525
ggt tat gee dee Gly Tyr Ala Pro 3530	teg etc ang mpg : Ser Lou Thr Leu / 3535	ərc tax aac agc gga Asn Tyr Asn Ser Gly 3540	acc ygl. eac: 31587 Thr Gly Asn

age eeg tte ggt ete ggt tgg gae tgt aan gto atg mea att egt egt Ser Pro Phe Gly Len Gly Top Asp Cys Asn Val Met Thr Ile Ary Arg 3545 3550 3555	31635
oge acc agt acc age gtg occ aat tat ga! gaw acc gat act tit otg Arg Thr Ser Thr Gly Val. Pro Asn Tyr Asp Ghu Thr Asp Thr Pro Leu 3560 3575	31683
gang oon goa ggit gan gitg tig gite gita give tiz eet gang gita ggit cas Gly Pro Glu Gly Glu Val keu Val Vol Ala Leu Ama Glu Ala Gly Glu 3580 3585 3590	31731
get gat aks nge agt gom tee tem tim eng ggs aks mat tig ggg mig Alm Amp Ile Ang Ser Glu Ser Ser Leu Clin Gly Ile Ami Leu Gly Net 3595 3600 3505	31779
ace the ace git ace ggt tat ego too ogt tig gaa age can tit age Thin The Thir Val Thin Gly Tyn Ang Sen Ang Lou Glu Sen His Phe Sen 3610 3615 3620	31827
egg tig gas tad igg Cam odd dam ada ada ggd gne add gat tid igg Arg Lau Glu Tyr Trp Gln Pro Gln Thr Thr Gly Ala Thr Amp Phe Trp 3625 3630 3635	31875
otg ata tac ago cor gar gga can gco cat tta ctg ggr ann ant cor Leo lle Tyr Ser Pro Asp Gly Gly Ala Nis Leo Leo Gly Lye Amb Pro 3640 3645 3650 3655	31923
can gon ogo all ago aat oos ota aal gut aac taa ara goy caa tgg Gln Als Arg Ile Ser Ago Pro Leu Aso Val Aso Gln Thr Ala Gln Trp 3660 3665 3670	31971
eta tig gaa ger tog gia tea ter ear gge gag tag att tat tat cag Leu Leu Glu Ala Ser Val Sor Ser His Gly Glu Gla IJr Tyr Glo 3675 3680 3685	32019
Lat oga god ges get ges not gat typ ges est ges gna etc ace god Tyr Arg Alo Glu Asp Glu Thr Asp Cys Glu Thr Asp Glu Leu Thr Alo 3690 3695	32067
යන පහසු කාල කලක මෙල ඉඩව එමල පසුව පසුව පසුකු ඉඩක ලැබ වන් ජනව ඉහුව His Pro Asn The The Val Gln Arg Tyr Lea Gln Val Val His Tyr Gly 3705 3710 3715	32115
ant ota acc god ago ges gta tit ood acg ota sel gga gmi gai oms Aen Leu Thr Als Sen Glu Val Phe Pro Thr Leu Asn Gly Asp Asp Pro 3720 3725 3730	32163
ide aan Let gge tog tig tie ligh tia gia tii gat tae ggh gag ege Leu Lya Ser Gly Trp Leu Pho Cya Leu Val Pho Acp Tyr Cly Glu Arg 3740 3750	32211
abs and ago the tot gas alg dog con the sas god acs agl and tog lys Asa Sec Leu Ser Glu Met Pro Pro Pho lys Ala Thr Ser Asa Trp 3755 3760 3765	32259
old tgo ogd 888 gad ogt til tod ogl lad gam tad ogt til gom tig Leu Cym Arg Lym Amp Ang Pho Ser Arg Tyr Glu Tyr Gly Pho Ala Leu 3770 3775 3780	32307
ogo acc ogg ogo lin tgt ogo csa ata otg sig lit neo ogt otg csa Azg Thr Arg Arg Leu Cys Arg Glu llo Lou Met Phe Ris Arg Leu Glu 3785 3790 3795	32355
acc etg tet ggt cag geo aaa gg: gwe gwt goa etc gea tiz gli. 1.cm	37403

Thir Lett Ser Gly Glin Ale Lyn: Gly Asp Asp Glu Pro Ala Lett Val Ser 3800 3805 3810 3815	
ogt otg a(a ctg get tat gec gen and gog gtg gtd agt adg otd gtt Arg Leu He Leu Asp Tyr Asp Glu Ash Ala Val Val Ser Thr Leu Val 3820 3825 3830	32451
tot gto uga egg gga cat gag caa gat ggo aca acg geg gto geo Ser Val Ang Ang Val Cly Ris Clu Gla Asp Cly Thr Thr Ala Val Ala 3835 3840 3845	32499
cty ccy cca tty gaa cty got tat cay cot ttt gan cca gan cam ama Leu Pro Pro Leu Glu Leu Ala Tyr Glu Pro She Glu Pro Glu Glu Lys 3850 3855 3860	32547
gea etc tym oga eta atg gal glâ etg geg aat the aac acc atc eaa Ala Leu Trp Arg Pro Met Asp Val Leu Ala Asn Fhe Asn Thr Ile Gho 3875 3870 3875	32595
ogo bgg caa obg oft gat obg caa ggo gaa ggo gta ooo ggt att obg Ang Typ Glo Leu Leu Amp Leu Glo Gly Glu Gly Val Pro Gly The Leu 3880 3885 3890 3895	32643
tat cag get ass sat ggg tgg tat cgs tot got cea cgt cag ace Tyr Gln Asp Lys Asn Gly Trp Trp Tyr Arg Ser Ala Gln Arg Gln Thr 3900 3905 3910	3 2 691
999 Saa gun aby aat gog gto doe tog ggo aaa atg caa cto ctt cet Gly Glu Glu Met Amu Ala Val Thr Trp Gly Lys Mes Glu Lou Leu Pro 3915 3920 3925	32739
ato and one get att com gat aan gen tea otg atg gat att est ggt. The Thr Pro Ala Ile Glin Asp Asn Ala Ser Leu Met Asp Ile Asn Gly 3930 3935 3940	32787
gat ggs cam dig gat tog gtt mic and ggt tog ggg cta agg ggt tal Amp Gly Clu leu Amp Trp Val lle Thr Gly Pro Gly Leu Ary Gly Tyr 3945 3950 3955	32835
car ago cag cat oca get ggc agt tgg aca cgt fill acg trig trig osc His Ser Gln His Pro Amp Gly Ser Trp Thr Arq Phe Thr Pro Leu His 3960 3965 3970 3975	32883
gur the mag ata gas het arm rat out ogn god cam oft god gat the Ala Leu Pro lle Glu Tyr Thr His Pro Arg Ala Glu leu Ala Asp Lou 3980 3985 3590	32931
ang ggg goo ggg cty frx: gal: bia ying old att ggt coo amma ago gtg Met Gly Ala Gly Leu Ser Amp Leu Vol Leu Ile Gly Pro Lym Ser Vol 3995 4000 4005	329 79
ngt tig Lai gee aat aan ngt gai ggt til ann gea gga ugg gai gig Ang Leu Tyr Ala Asm Asm Ang Asp Gly Phe Tir Glu Gly Ang Asp Val 4010 4020	33027
guy raa ten ggt. ggt. att een etg oog tia oog nge gee gat geg egt. Val Glm Ser Gly Gly Ile Thr Leu Pro Leu Pro Gly Ala Asp Ala Arg 4025 4030 4035	33075
aag tta gtg god til agd gad gta etd ggt tea ggd daa gea dal 1929 Nys Leu Val Ala Phe Ser Asp Val Tex Gly Ser Gly Cln Ala His Leu 4040 4045 4055	33123
gtt gaa git agt gog aog aaa gid soc too tge coa aat oig gga oat Val Glo Val Ser Ala Thr Lys Val Thr Cys Trp Pro Asn Leu Gly His	33171

1060	4065	4070
	e ato ace thig may gge thi The Tim Leu Pro Gly the 4080	
gun aal thi aat ooi gat Ala Asm Min Aun Puo Aug 4090	t cga gitt cat etg gen gal o Ary Val llie Leu Ala Asp 4095	olg gad ggt agt. 33267 Deu Asp Gly Ser 4100
ggt cot goo gat otg att Gly Pro Ala Asp Leu Ile 4105	t tat gil cel gol gar cac e Tyr Val His Als Asp Bis 413.0 4135	Lou Asp Ile Phe
ago aat gae 29t 99t 2ac Ser Ash Gin Ser Gly Aer 6120 4125	c ggt ttt gca caa cca tto n Gly Phe Ala Chn Pro Phy n 4130	acalete ogtittt 33363 The Len Arg Phe 4135
ect gae gyn eig ogb 600 Pro Amp Gly Leu Arg Pha 4140	l gat gat act tyc cag cta 2 Asp Asp Thr Cys Gln Leu 4145	cea gtg gct gat 33 4 11 Gln Val Ala Asp 4150
	t gic ago oig ale oig ago 1 Val Sor Leu The Teah Ser 4160	
Ala Pro His His Trp, Art 4170	r tgo gat otg acc zac gog g Cys Aep Leo Thr Aen Ala 6175	lys Pro Trp Leu 4180
Cld Byt gas atg aac aac Leu Sor Glu Met Asa Asa 4185	c aac atg goa goo cat cac hAan Met Gly Ala His His 4190 4195	The Leu Ris Tyr
Arg Ser Ser Val Glo Phe 4200 4209		Ala Ala Leo Ala 4215
acc ggs cas ace ong gto The Chy Gin The Pro Vsl 4220	c tgt two otg ood tto oog 1 Cyn Tyr Leu Pro Phe Pro 4225	gto cat acc ctg 33651 Val His Thr Len 4230
ੀਬਰ Cla ਸੋਖਟ Glu ਤੀਖ਼ਟ Gli 4235	I gat gam ate age gge am I Anp Glu Ile Ser Gly Arm 4240	1 केट्ट (42) Vel Thr 4245
act tha ogt tac got tac Thr Teu Ang Tyr Ala His 4250	r god god togg god gga ogt s Gly Ala Toop Assp Gly Arg 4255	gag togg gaz tit 33747 Ghu Ang Ghu Phe 4260
ege age tit ggo tat gti Amg Gly Phe Gly Tyr Val 4265	t gag cag aca gac agc cat l Glu Glu Thr Asp Ser His 4270 4275	: Gin Leu Ala Gin
ggc aat nom com naa egi Gly Ami Ala Pro Glu Am 4280 4280	t aca tea teg gea ett ace g Thr 9ec Pro Ale Leu Thr 6 — 4290	e aaa aac tgg tat 33843 Hys Aso Tro Tyr 4295
give acc qua ato cot gad Ala Thr Cly The Pro Ch 4300	g gta ged aat acg dta tot u Vel Asp Asm Thu Lou Ser 4305	goo ggg tat tgg 33891 Ala Gly Tyr Trp 4310
ege gat gat aog eag get Ary Gly Asp Thom Glo Ala 4305	t the act ggt thi acg em a Phe Thir Gly Phe Thir Pro 4320	can tit act etc 33939 His Phe Thr Leu 4325

tgg maa gag Trp lys Glo 4 330	Gly Ign: Asp	gtt oca elg Val Pro Leu 4335	Thr Pro Glu A	pat gae eac aat Nap Aap Ris Asa 140	33987
ote lac teg Lac Tyr Typ 4345	Ikm Irm Img	gca cto amb Ala Leu lys 4350	ggt caa cca c Gly Gln Pro I 4355	rtg ogt agt gaa æv Arg Ser Glu	34035
che tae ggg Leu Tyr Oly 4360	eta gat ggc Leu Asp Gly 4365	ago gua rag Ser Ala Glm	cag aag atc o Glu lys Ile I 4370	rc tat aca gtg To Tyr Thr Val 4375	34083
act gaa toc Thr Glu Ser	ngo oma caa Ang Pro Gla 4380	Val Arg Gin	tta caa gat a Leu Gìn Asp A 1385	we act acc ctt wan Thur Thur Leu 4390	34131
Ser Pro Val	cte tgg gee Leu Tip Alm 4395	tca gtg gtg Ser Val Vel 4400	gam agt ogt a Glu Ser Arg S	egt tat can tat Eer Tyr His Tyr 4405	34179
gee out atc Glu Arg Ile 4410	Ile Sex Arap	eee caa tge Pro Glo Cys 4415	Aen Gir Aep 1	ato act otg too Ne Thr Leu Ser 120	34227
agt gad ota Ser Asp Leu 4425	tto ggg coa Phe Gly Gln	cce ctg aaa Pro Leu Lys 4430	çay yiti tes ç Glin Val Ser \ 4435	ytg caa tat con Yal Glo Tyr Pro	34275
oge oge aat Arg Arg Asm 4440	aaa eea aca 1ys Pro Tho 4445	Thin Assa Pro	tan oxo gel a Tyr Pro Asp 1 4450	wa cha cce gat Mr Leu Pro Asp 4455	34323
act otg ttt The Leu Phe	goo ago agt Ala Ser Ser 4460	Тук Авр Авр	നുടെ cea cea d Un Un Un I 4465	th tty ogg the Jou lou Arg Leu 4470	34373
Thu Tyr Glo	icaa tee agt iGtn Sem Sem 4475	tgy cat cat Trp Ris His 4480	cta att gct a Teu lle Ala)	aat gaa ete aga Van Glu Leu Arg 4485	34419
gtg tta gg a Val Leu Gly 4 49 0	'Ikou Pro Ass	ggt aca cgc Gly Thr Arg 4495	Ser Asp Ale 1	tte act tae gat Phe Thu Tyr Asp 500	34467
get amm cac Als Sys Him 4505	:gtg oct gtt :Val Pro Vel	çat ggt tta λφρ Gly Leu 4510	aat otg gna g Asn Leo Glo / 451.5	got ota tgt get Ala Leu Cye Ala	34515
gma sat agn Glu Amn Ser 4520	: ctg All. gcz leu lle Ala 4525	asp asp lys	endi ogo gaa i Pro Arm Glu i 4530	tac oto eec meg Nyr Leu Asn Glin 4535	34563
cas ega seq Gln Arm, Thu	tto tat acc The Tyr Thu 4540	· Asp Gly Lys	асс дат дда а Thr Ажд Gly (48 4 5	aan aat oom bog Iys Aso Pro Thr 4550	34611
ome otg aaa Pro Leu lys	aca cog aca The Pro The 4555	oga cag get Arg Gln Alo 4960	Low He Als 1	ttt acc gam acg Phe Thr Glu Thr 4565	34659
geg gta tta A)ə Val Lev	non one to	ete tta tee	gea hitt gat g	ജ്യ: പുല്ലി, ക്ഷീൻ ക്രൂട്ട	34707

one get gas the con ggs: off obg ech can got ggs had can see gan the Asp Glu Leu Pro Gly Leu Leu Thr Glu Ale Gly Tyr Glu Glu Glu 4585 4590 4595	34755
oot, tak ong the east east east east east east east eas	34603
ass ago tak ach yet too gga act gag gtA (sa tit tgg ogt oot gio Lys Gly Tyr Thr Asp Tyr Gly Thr Glu Val Gln Phe Trp Arg Pro Val 4620 4625 4630	34851
gos caa ogt aac acc cag 13a acc ggg aan acg act ots aan tog gat Ala Gln Arg Ash The Cln Leu The Gly Lys The The Leu Lys Tep Asp 4635 4640 4645	34899
acc cac tac tgt gtc atc act (00% acc cas gac gcg gct ggt ttg act Thr His Tyr Cys Val lie Thr Gin Thr Gin Asp Ala Ala Gly Len Thr 4650 4655 4660	34947
gte tea ger aat tat gar tog ogt tit etc acs eet sig Owa eig act Val Ser Ala Ash Tyr Asp Tip Arg Phe Len Thr Pro Met Gin Len Thr 4665 4670 4675	34995
gat atu san gan sat gtg cat atu ata ann ttg gat gog cta ggs ogc Asp lle Asm Asp Asm Val His lle lle Thr Leu Asp Als (eu Gly Arg 4880 4685 4690 4695	35043
cet gite act caa egi tit teg goa ate gaa ast ggi gitg gea acs ggi. Pro Val Thr Gin Arg Phe Trp Gly Tin Glu Asn Gly Val Ala Thr Gly 4700 4705 4710	35091
tac tot top oce gram gos eas now toe act oce ose gho get gite amit Tyr Ser Ser Pro Glu Ale Lym Pro Ele Thr Pro Ino Val Asp Val Ash 4715 4720 4725	35139
got god All you dig acc gga com old oot gto gog dag tgi dig gto Ala Ala Ile Ala Leu Thr Gly Pro Leu Pro Val Ala (In Cys Leu Val 4730 4735 4740	35187
Law goy tong gax; agt tong atgreeg often the gift day gas acc the acc Tyr Ala Pro Asp Sex Try Met Pro Lew Phe Gly Ohn Glu Thr Phe Asm 4745 4750 4755	35235
aca tha any may gaa may caa aag ama ctn egt gat tha may ath atc Thr Leu Thr Ghu Ghu Ghu Lys Thr Leu Arg Asp Leu Arg Ile Ile 4750 4765 4770 4775	35283
ace gas get tog ogt att tgo gos oly got ogo ogt tgg ota coa Thr Glu Asp Trp Arg The Cyo Ala Leu Ala Arg Arg Arg Trp Leu Gln 4780 4785 4790	35331
agt cas ass got age are one the att and one the act as and age atc Ser Glu Lys Ala Gly Thr Pro Leu Val Lys Lou Leu Thr Ash Ser Tlo 4795 4800 4805	35379
ggt tta oot ooc cae AA; oto atg otg get ang gan ogt tat gad ogt Gly Lau Pro Pro His Amn Lau Mel. Lou Ala "Mrr Amp Ang Tyr Amp Ang 4810 4815 4820	35427
gat tot gaz cog caa att ogt com caa gto gom old ogt gat ggt tit Asp Ser Ghu Gho Gho The Arg Gho Gho Val Alo Phe Ser Asp Gly Phe 4825 4830 4835	35475
AND COST. THE THE COST SET SET SEE COST COST SEED SET	35523

Ciy Arg Leo Lou Gin Alb Alb Val Arg His Glu Ala Giy Glu Ala Trp 4840 4845 4850 4855	
can ogt sac caa gac ggt tot otg gtg aco aaa atg gma got aco aaa Glin Arg Asn Glin Asp Gly Ser Leu Val Thr Lys Met Glu Asp Thr Lys 4860 4865 4870	35571
seg ege tgg geg att aeg gga ege aet gaa tat goe aat aag ggg eag Thr Arg Tro Als Ile Thr Gly Arg Thr Glo Tyr Asp Asm lys Gly Gln 4875 4880 4885	35619
gog eta oga aki fak kag kom hat blo okk aat gad lag oga tat gog Ala Ile Arg Tho Tyr Glo Pro Tyr Phe leu Aso Asp Tro Arg Tyr Val 4890 4895 4900	35867
agt gat 990 890 gov ags ees gag gov tat gov gat ent cat ato tot Sex Amp Amp Ser Ala Ang lys Glu Ala Tyr Ala Amp Tor Him Ile Tyr 4905 4910 4915	35715
gat cog att ggg Cgg god atc coa gtt atc acg gca aaa ggc tgg ctg Amp Pro Ile Gly Arg Glu Ile Glu Val Ile Thr Ala Dym Gly Trp Leu 4920 4935 4930 4935	357 6 3
egg cag aar caa tat tte eeg igg tit see gtg agt gaa gat gaa aat Arg Glu Aso Glu Tyr Hos Pro Tro Pha Thr Val Ser Glu Aso Glu Aso 4940 4945 4950	35811
gat tig toe get gae geg ete gig taa tigaateaag attegetegi. Aap Leu Ser Ala Asp Ala Lou Vel 4955 4960	35858
ttaatgttaa ogagogaata taatatacct aatagatttc gagttgcagc goggoggcaa	35918
gtgaacgaat receaggage atagataact atgtgactgg ggtgagtgaa ageagemaan	35 97 B
asagraycag chigasagat gangggtata aatangaano tgcattgtga gttotaaata	36038
gagtagoago ataittisti goottikusti, koskagytaa taaaattooo tigotgiaaa	36098
aatotgtoat catgagaact aaaaalaaca achittofoti, chighaagaga aatomatmat	36158
tosattaaas atgitataga atrigaatoa agenratikg tiggotoeto vaaaatataa	36218
acateogest egytaataas agetgatgen askayaasti lektitittat oosaagtgae	36278
atatyteest setematace agastastta gatatacema ancembilam etagtameet	36338 '
eattgatatt ttaeattact ttteetatas egetgaetta aattaateme Autmonteex	36398
gtgatgaaat tataaaagtt aacattatoo gatagataas aaccatgotg Uugomaatta	36458
aatoggetet titeteeett tittataassa tiaacestic etiilluusse eitatitaes	36518
cragosatac tigasagass agrogotitis congenicat taxetecoge astacogniti	36578
ней/живети: geneal/www.attgactoct generaliteg tottaccatt aataacaacc	36638
inisaconas Laacilkaag catematago cootaasaat aacgtassaa agaasatasc	36698
normandata ataatititeg tylkutgegt teteanoaga gasatagaag amacamtaat	36758
agaaagaaaa gootaagata aaaatataat dacaggaaaa gayttaxoax caagaaagco	36818
Assastassa anacemaco satassasas casacesata contastina asacesatat	36878

tttoograda gataaasagi logadaaata tgaaagataa tttatttoom tatatgatag 36938 attatament amcamcatgo stale/ptam amcamcatg gentatotta atgatatota 36998 atcogocitky titygggatti grugawwyyk artittoscat aataqatata aaaqoagaac 37058 agotakhgan ataataagga (egagathi), attittatta aaacaataac gaagatteat 37118 tataleeggo aatgazaasa emudigatga saataattit ttattiotat taattatata 37170 exelugatgig abouttrees. Afestatosa igotaciasi gganiosota sigiossasai 37238 caaatoatat makaiicean teetgaatga tgoogroonga agaaagaaka cagcaacaal 37298 esassanatgo assessotta attomastan pominatore attacegose asystactet. 37358 сававляли: Acagatgasa ggtaatgcan atmattames thlooghees езиамостат 37418 aangaayasa ataactatog gabangozot atxaatesan saaangalan gantaasaa 37478 қажықтітін itaoctanon амулямоды. getigestic icciitycey eaggeeasse 37538 cottatgita atomanisma almoviatata teoristiama gistetgoping temmatemaa 37598 tgattitzig (sgccstcig gastaataat attggaagat aaagittatta aaaccicaas 37658 gataccacty wouldbycop yasytestas aagaaaasyg satataatys cattititati 37718 eccagacypa cattlettta teutametit atattecaag gesteagega tiattaaatt 37778 retartgeet etetamasen amambelesa taatgteett getgaatett tagggaatet 37838 ogtootggma tgczwakata aatsgttact gaaaacaata cattgattit taattasata 37898 37948 otgoogaliah gacottaatg atgotacttt attitooogt attomating

<210> 12

<211> 954

212> PRT

<213> Photorhabdus luminescens

<400> 12

Met Lym Ach The Amp Pro Dys Leu Tyr Gin Lys Thr Pro Val Val Amn 1 10 1.5

The Tyr Asp Asp Ang Gly Len Thr The Arg Asp The Asp Phe His Arg 20 25 30

The The Ala Asia Gly Asia The Asia II.e Arg II.e The Arg Ris Glib Tyr 35 40 45

Asp Ser Lea City His Lea Ser Gin Sor Thr Aup Pro Arg Law Tyr Gia 50 60

Also Lys C(n Lys: Ser New Piet Leu Trp Gln Tyr Asp Leu Thr Gly Ass 65 70 75 80

the Leu Cys Thr Glu Sor Vol Asp Als City And Tho Val Thr Leu Arab 95 90 95

Asp Tie Chu Gly Arm Pro Leo Leo Thr Val Thr Ala Thr Cly Val Tie 100 105 110

Gin Thr Arg Gln Tyr Glu Thr Ser Ser Leu Pro Gly Arg Leu Leu Ser

		115					120					125			
Λ9Ţ	Thr 130	Glu	Gin	Ile		Glu 135	Ľγs	יולני	Ser	Arg	Ile 140	Thr	Glu	Arış	Lou
11¢ 145	נלטינו	λla	Gly	A en	5er 150	Glu	Ala	Glu	[ge;	A(a) 155	His	Acm	īæs	λla	883 160
Gln	Cys	Val	Arg	н <u>і</u> я 165	туг	asp	Thr	Ala	Gly 170	Və.1	ДЛ.	Arg	teu	Glu 175	Ser
Leu	Ser	Leu	Thr 180	Cly	TI or	Val	Leu	Se r 185	GLn	Ser	Ser	Gln	Leu 190	Leu	Se1
Asp	'thur	G) n 195	Glu	Λlε	Ser	Ттр	Thr 200	Gly	Asp	Aem	Glu	'Itur 205	Val	Trp	Gln
A;त)	16et 210	स्या	Ala	Asp	ASTO	Ile 215	Tyr	Тπ	Thur	تجها	8en 220	Ala	She	Mego	AlA
Thr 225	C) y	Ale	Leu	Leu	Tîm 230	ദ്ധ	Thr	ABD	Ala	18/5 235	Gl _Y	æn	Ile	Gln	Arg 240
pen	Thr	Tyr	Жр	Val 245	УĴЭ	Gly	GJI	Leu	ABO 250	Gly	Ser	Trp	Leu	Tha 255	Leu
Lys	Perp	Gln	Pro 260	Glu	Chi	Val	Лe	TJ.0 265	Arg	Sear	[es.)	Thr	Туг 270	ær	Ala
Ale	Gly	Gln 275	-	Leu	Arg	Glu	Glu 280	His	Gly	Asn	Gly	Val 285	1le	Tîtur	Glu
Тус	Ser 290		Glu	Pro	GJu	Thr 295	glu	glu	ren	Ile	300 CJÀ		Lys	Thr	His
Ary 305		· Ser	Авр	УŢЭ	1ys 310		Leu	Gju	Asp	1eu 315	-	тут	Glu	Тух	As p 320
				325					330	•				335	Алц
Phe	Trp:	HŻE	340		nys	vs)	Ala	Pro 345) Aen	Tîrx	Tyr	Thi: 350	_	N EQ
Ser	Leu	7yr 359		Lou	ılle	Ser	360		· Gly	, yrd	: Glu	Met 365		Asn	lle
G ì y	Gln 370) Ser	AET.	Gln	1 .e 0 375		Sec	Leu	TÎTE	Leu 3MD		Seor	Asp	A.am
Aso 385		· TYI	The	Near.	390 1AL		. Arc	Tre	Туг	· Thr 395		· Asp	Arg	Gby	Gly 400
le pr	ie:	Thu	Tye	11e 405		llís	: Ser	Ser	Pro 410		o Tilor	. Ch	AEJ:	1 ASI: 415	тут
Tîtu	Tìu	. Asc	11e 420		· Val	. Ser	ABC	425		. Asr	ALTO	ı Ala	430		seo
'Itu	Teu.	1 Thu 43:		ı Neg	e Pro	ALE	G(1 44(LАнд	ι Νέε	i lei	. Pive 449		Ala	Gly
Gly	/ Hi: 450		, ya	13 10	i Let	11e 459		- Gly	/ Gla	n A sar	1 Lou 460) Til) Afir	ימינדו מ

Arg Oly Clu 465	ිනා Cln His 470			Lye Arg . 475	Aesp lym (Cly Al a 480
Aen Aep Aep	Arg Glu Trp 485	Tyr Arg T	Ут <i>S</i> ст : 490	ිපැ <i>වන</i> ු ඉ		Arg Ile 495
len lys Ilo	Asn Glu Glo 500		er Ser. 305	Man Sec	Cla Thr 510	Gln Avg
ile Thr Tyr 515	Leu Pro Ser	ieu Glu I. 520	ਲਗ ਮੈਟਜੂ ਹੈ		Gla A නා 525	Ser Than
12e Thr Thr 530	Glu Asp Leu	Gln Val (535	The Thir	Val Gly 540	Glu Ala	Gly Arg
Ala Glm Val 545	Arg Val Lou 550			Gly Gla 555	Pro Glu	Asp Ile 560
Asp Asn Asn	Glm Leu Arg 565	Tyr Ser T	lyr Asp . 570	ABN Leu	Ile Gly	Ser Ser 575
Gln Leu Glu	Leu Asp Scr 580		Slu Ile 185	lle Ser	Olu Glu 590	Glu Tyr
Tyr Pro Tyr 595	Gly Cly Thr	Ala Leu 1 600	Prp Ala	The Arg	PAS VAC	Than Glu
Ala Ser Tyr 610	Lys Thr Ile	: Ang Tyr 3 615	Ser Gly	Lys Glu 620	Arg Asp	Ala Tur
Gly Leu Tyr 625	Tym Tyr Gly 630			Cln Pro 635	'Irp Val	Gly Arg 640
Tap Lea Ser	Ala Asp Pro 645	Ale Gly 1	The Val 650	Asp Gly	Leu Ası	Lou Tyr 655
Arg Met Val	Arg aso as 660		Thr Le u 665	Leu Asp	Pz() Asp 670	Glγ Leu
Met Pro Thr 675	lle Ala Glo	Arg lle A 680	Ala Ala	Less Glin	Lys Asn 685	lys Val
Ala Asy Sur 630	Ala Pro Ser	ena Mir A 695	Asn Ala	The Ago 700	Val Ala	Ile Asn
Ile Arg Pro 705	Pro Val እነሪ 710		Pro Thr	les: Рхо 715	Lys Ala	Ser Tho 720
Ser Ser Gin	1 Sec. That Tha 725	Tyr Pro I	11e 1ys 730	Ser Ala	Ser Ile	lys Pro 735
The The Sec	'Gly ଜଣ' ଚଥା 740		Ala Pro 745	l en a Sur	liro Val 750	Gly Assi
Tys Sor Tha 755	Pro Glu Ile	: Ser Leu I 760	Pro Glu	Ser That	Glm Ser 765	Aso Ser
Ser Ser Ala 770	The Ser Th	775 Aen Lenik	Gln l ys	Lys Ser 780	Phe Thir	Courtyr (
Arg Ala Aაგ 785) Asn Arg Se 1790		Asp Met	Gln Ser 795	lys: Hoe	ETO Glu HQO
Gly Phe Lys	ะ Ala Trp ฟีข 805	r Pro Leu I	Asp Tta 810	Lyca Net	Ala Arg	Gln Phe 815

Ala Ser Val Phe Ile Gly Glu Lya Asp Thr Ser Aso ເວດ Pro Lys ເປັນ 820 825 830

Thu Val Lym Asn The Asn Thu Trp Gây Thu Lym Pro 1993 Lew Asn Asp 835 840 845

Low Ser Thr Tyr Ile Lye Tyr Thr Lye Asp Lye Sar Thr Val Trp Val 850 $\,$ 860 $\,$

Ser Tim Ala Tie Asi Tim Glu Ala Gly Gly Gln Ser Ser Gly Ala Pro 865 870 875 880

Leu His Glu Ils Asn Met Asp Leu Tyr Glu She Thr Ils Asp Gly Gln 885 890 895

bys Len han Pro Leu Pro Arg Gly Arg Ser Lys hap arg Val Pro Sar 900 905 910

led led for Amp Thr Pro Clu Ile Glu Thr Ala Ser Ile 1le Ala Led 915 920 925

Ash His Gly Pro Val Ash Asp Ala Glu Val Sor Phe Leu Thr Thr 1,20 930 940

Pro Leu Lys Asn Val Lys Pro Tyr: Lys Arg 945 950

<210⊳ 13

<211> 2522

<212> FRT

<213> Photorhabdus luminosogra

<**400**~ 13.

Wet The Leu Lys Gly The Asm Met Asm Ser Pro Vel Lys Glu The Pro 1 5 10 35

Asp Val Lau Lys The Ghn Cys Gly Phe Ghn Cys Leu Thr Asp The Sec20 25 30

His Ser Ser Phe Asn Olu Phe His Gln Gln Val Ser Glu His Leu Ser 35 40

Trp Ser Glu Ala His Asp Leu Tyr Ris Asp Ala Gln Gln Ale Gln Lys 50 55 60

Asp Asm Arg Leu Tyr Giu Ala Arg Ile Leu Lys Arg Thr Asm Pro Gin 65 70 75 80

Leu Gin Astr Ala Val His Leu Ala Ile Val Ala Pro Am Ala Glu Leu 85 90 95

Ilo Gly Tyr Aso Aso Glo the Ser Gly Arg Ala Ser Glo Tyr Val Ala 100 105 110

Pro Gly Thr Val Ser Ser Met the Ser Pro Alz Ala Tyr Lou Thr Glu 115 120 125

len Tyr Arg Gio Ala Arg Asn Leo Hie Ala Ser Asp Ser Val Tyr Arg 130 135 140

Leu Asp Thr Arg Arg tro Asp Leu Lys Ser Met Ala Leu Ser Gln Gln 145 150 155 Aen Met Asp Thr Glu Leu Ser Thr Leu Ser Leu Ser Asn Glu Leu Leu 2.70 Ion Glu Ser Ile Lys Thr Glo Ser Lys Leu Asp Asn Tyr Thr Glo Val 1B5 Met Glu Met Lou Ser Ala Phe Arg Pro Ser Gly Ala Thr Pro Tyr His Asp Ala Tyr Glu Asm Val Arg Lys Val Ile Gln Les Cln Asp Pro Gly leo Glo Gln Leo Asn Ala Ser Pro Ala Tle Ala Gly Leo Met His Gln Ala Ser Leu Leu Gly The Asm Ala Ser The Ser Pro Clu Lou The Asm Lie len Tim Giu Giu Ile Tim Clu Cly Aso Ala Giu Giu Leu Tyr Lys Dys Asn Hoc Cly Asn the Glu Pro Ala Ser Leu Ala Mei. Pro Glu Tyr Leu Arg Arg Tyr Tyr Asn Leu Ser Asp Glu Glu Leu Sur Gln Phe 13e 295 Gly Lye Ala Ser Asm Phy Gly Gln Gln Glu Tyr Ser Asm Asm Oln Leo 315 Ilo Thr Pro Ile Val Asm Ser Asm Asp Gly Thr Val Lys Val Tyr Arg lle Thr Ary Glu Tyr Thr Thr Ash Ala Ash Glin Val Amp Val Glu Lea Phe Pro Tyr Cly Gly Glu Asin Tyr Clin Lou Asin Tyr Lys Phe Lys Asin 365 360 Ser Arg Glin Asp Val Sir Tyr Leu Ser Ile Lys Leu Asn Asp lys Arg Clu Lou He Arm the Glu Gly Ala Pro Cln Val Asa the Glu Tyr Ser Glu Ris Ile Thr Leo Sor Thr Thr Asp Ile Ser Glo Pro Pho Glu Ile Gly Leu Tun Ang Val Tyr Pro Ser Ser Ser Top Ala Tyr Ala Ala Ala 425 Lye Phe Thr Ile Glu Glu Tyr Agn Gln Tyr Sex Phe Lou Leu Lye Leu Asm Lys Ala lle Arg Leo Sor Arg Ala Thr Glu Leo Ser Pro Thr lle Ican Clu Ser Ile Val Arg Ser Val Aon Clin Glin Leu Asp ile Ash Ala

Clu Val Leu Gly Lys Val the Leu Thr Lys Tyr Tyr Met. Glo Arg Tyr

Als The Asm Aka Chu Thr Ala Leu Ile Leu Cys Asm Alm Leu Ile Ser 500 505 510 Glin Arg Ser Tyr Asp Asn Glin Pro Ser Clin Pho Asp Arg Leu Phe Asn 515 520 525 The Pro Lou Leu Ash Gly Gln Tyr Phe Ser Thr. Gly Asp Glu Glu Ile Asp Leu Asm Pro Gly Ser Thr Gly Asp Trp Arg Lys Ser Val Leu Lys Arg Ala She Amn Ile Amp Amp Ile Ser Leu Tyr Arg Leu Teat Lys Ilo 565 570 The Ash Ris Ash Ash Gin Ash Ghy Lys Ite Lys Ash Ash Leo Ash Ash Let Ser Asp Let Tyr The Gly Lys Let Let Ala Ghi The His Ghi Loo Thr Ile Asp Glu Leu Asp Leo Leo Vel Ala Val Gly Glu Gly Glu 610 615 The Asm Ten Ser Ale Lie Sor Asp bys Gin Leo Ala Ala Leo Ile Arg LyB Leu Ash Thr Ile Thr Val Trp Leu Gln Thr Gln Lys Trp Scr Alzi 645 650 650 Fire Glin Leu Phe Val Wet Thr Ser Thr Ser Tyr Asm tys Thr Leu Thr 660 665 670 Pro Glu Ile lys Asn Leu Leu Asp Thr Val Tyr His Gly Leu Gln Gly Phe Asp Lys Asp Lys Ala Asm Leu Heu His Val Mot Ala Pro Tyr Ile Ala Ala Thr Leu Glin Leu Ser Ber Gliu Ason Val Ala His Ser Val Leu 705 710 715 720 Low Trp Ala Asp Lys Leo Lys Pro Gly Asp Gly Ala Ket Thr Ala Glu 725 730 735Lye Phe Trp Asp Trp Len Asn Thr Glm Tyr Thr Pro Asp Ser Ser Glu 740 745 750 Val Leu Ala Thr Gin Clu His the Val Gin Tyr Cys Gin Ala اها، Ala Gln Lew Glu Mot Val Tyr His Ser Thr Gly Xle Asm Clu Asm Alo The 770 775 780 Arg Lou Sto Val Thr Lys Pro Glu Met Phe Gly Ser Ser Thr Glu Ala Val Pro Ala Bis Asp Alm Leu Ser Leu Ile Met Leu Thu Arg Phe Ala App Trp Val Ash Ala Leu Gly Glu Lys Ala Ser Ser Val Lou Ala Ala B25 Phe Glu Ala Asn Ser Leu Thr Ala Clu Cln Lau Ala Asp Ala Met Asn

ion Asp Ala Asm Leu Leu Lou Glm Ala Ser Thr Glm Ala Glm Asm His

	850					855					860				
Gln 865	His]सा	Pro	Pro	Va 1 870	Ħìr	Gln	Lys	A sn	Ala 875	Phe	වසර	Cys:	OLS.	Ulu H80)
Sear	lle	λер	7fm	11e 885	Leu	сĵи	TIP	Val	Aen 890	Val	Ala	GΓυ	Gln	£சுப 895	ley i
Va).	Ala	Pro	ദ്വമ 900	Gly	Val	Sec	Мa	Len 905	Vel	Giy	le'a I	A:T	Tyx 310	Ile	Glv
Pen	A a n	3) n 915	Lyes	Jles	Pro	They	Тут 920	Ala	Cln	Ттр	Glu	Ser 925	Ala	GΙγ	GľΩ
Ile	Lец 930	Thu	Ala	Gły	Leu	Asn 935	Ser	GJU	gju	Мa	ле р 940	Ile	Len	Ai G	Ala
Ptx: 945	Leo	Asp	сул	Ser	Лгу 950	Sex	Al&	Ala	Lou	955 955	Thu-	Тут	ፓ ነ፲	Ile	Arg 960
GLn	Val	Ala	Lyb	Pro 965	Ala	Ala	Ala	ll€	Ly18 970	Ser	Y rd	Asp	ж	IД; 975	Tyr
Gľu	Тут	leŭ	980 080	11e	Asp	Asto	Gln	Val .985	Ser	Ala	Ala	lle	Σγ:: 9 9 0	The	Thu
Arg	Ile	Ala 995	Glu	Ala	Ile		\$:: т′ 1000	IJe	G l n	Lou		Val 1005	Asn	Arg	TÍM
Lou :	G10 1010	ASD	Val	61 0	Gչս :	Aகா 1015	Ala	هنظ	Ser		Val 1020	lle	Ser	Arg	Gln
Phe 025	lthe	IJe	Asp	Trp	АЭР 1030	Γ¥ε	Тух	æn		Arg 1035	Тух	Ser	ጥነፈተ		Ala 1040
GΙA	Væ1	Ser	gln	1045 1045	val	Tyr	TYI		Glu 1050	Λsn	Tyr	Πc		1270 1055	Tha
			1060 GJA					1065				•	(070		
		1075	Gln				1080				-	LDB5			
Tyr	Leu 1090	Tìr	Ser.	Phie	Glu	Gl.n 1 09 5	Ve 1	Alm	Àπn	Leu :	Lys L100	Val	Ile	Sett.	κLΛ
тул 105	His	Æp	Asin	Tile:	Asn L110	ÀFO	ASP	Gln		Le u 1115	The	Tyr	Phé		Gly 1120
Leu	Ser	Glu	"Shar S	Asp 1125	Thr	GJA	Glu		тут 1130		Arg	Ser		Авр 1135	Dis
Sen	lye	Phe :	Ser 2140	æp	લોપ્ર	lys		هلد 1145	Aža	Aran	Ala		Ser 1350	Glu	Trp
(Ais:	1.ys:	11e 1155	<i>শু</i> শু>	Cyrs	Pro		ASD 1160	Pro	Ίንፐ	Arg		Tim 1165	Ile	Ατy	Pro
Va2 :	M at 1170	ጥ _ያ ን	lys	ජනා	Arg	Leu 1175	Tyr	Leu	Leu		Leu 1180	Glγ	Cln	Γλε	Glu
Ile 185	Tìu	LyB	Gln	Thu:	61y 1190	Asır	Ser	Uys		Gly 1195	lyr	Gln	Thr		ጥ 亚 1200

- Amp Tyr Arg Tyr Clu Leu Lys Leu Ala His Lite Arg Tyr Amp Gly Thr 1205 1220 1215
- Trop Asm Thur Pro Ile Thur Phe Amp Val Amm Glo Liye Ile Sex Lys Len 1220 1225 1230
- Glu Leu Ala Lys Ash Lys Ale Pro Gly Lnu Tyr Cys Ala Gly Tyr Gln 1235 1240 1245
- Gly Glu Asp Thy Lew Lew Val Wet Phy Tyr Asm Gln Gln Amp Thr Lew 1250 1255 1260
- Asp Ser Tyr Lys Thr Als Ser Met Gln Gly Leu Tyr Ile $9h\pm$ Als Asp 265 1270 1275 1280
- Met Glu Tyr Lys Asp Met Thr Asp Gly Gln Tyr Lys Ser Tyr Arg Asp 1285 1290 1295
- Asn Ser Tyr Lye Glin Phe Amp Thr Amn Ser Val Arg Arg Val Asn 1300 1305 1310
- Arg Tyr Alm Glu Asp Tyr Glu Ile Rno Ser Ser Vel Asm Sor Arg Lys 1315 1320 1325
- Gly Tyr Asp Trp Gly Asp Tyr Tyr Leu Ser Met Val Tyr Asm Gly Asp 1830 1835 1840
- The Pro Thr Ile Ser Tyr Lys Als Thr Ser Ser Way Low Lys Ile Tyr 345 1350 1355 1360
- The Sor Pro Lys Lea Arg The The His Asm Gly Tyr Glu Gly Glm Glm 1365 1370 1,375
- Arm Asn Gln Cys Asn Lau Met Asn Lye Tyr Gly Lye Leu Gly Asp Lye 1380 1385),390
- Pie fle Val Tyr Thr Ser Leu Gly Val Ash Pho Ash Ash Ser Ser Ash 1395 1400 1405
- Tys Leu Met Phe Tyr Pro Val Tyr Gln Tyr Asn Gly Amn Val Ser Gly 1410 1415 1420
- Lem Scr Cln Cly Arg Lem Lem Phe His Arg Asp Thr Ash Tyr Ser Ser 425 1430 1435 1440
- Lys Val Glu Ala Trp Ile Pro Gly Ala Gly Arg Ser [co Thr Asm Pro 1445 1450 1455
- Ast. Alo Ale IIe Gly Asp Asp Tyr Alo Thr Asp Ser Lea Asp Lye Pro 1460 1465 1470
- Ash Asp Leu Lys Glu Tyr Val Tyr Met Thr Asp Ser Lys Gly Thr Ala 1475 1485 1485
- Thu Asp Val Ser Gly Pro Val Asp The Asm Thr Ala Jie Sam Pro Ala 1490 1500
- Lys Val Gln Val Thm Val Lys Ala Gly Ser Lys Glu Gln Thr Phe Thr 505 1510 1515 1520
- Ala Asp Lya Asm Val Ser Ile Glo Pro Ser Pro Ser Phe Asp (Nu Mot 1525 1530 1535
- Asm Tyr Gln Pha Ash Ala Leu Glu Fle Asp Gly Ser Ser Leu Asm Phe 1540 1545 1550

- The Asn Asn Ser Ala Ser Ile Asp Ile The Phe The Ala She Ala Glu 1555 1560 1565
- Asp Gly Arg Lys Leu Gly Tyr Glu Ser the Ser The ν ro The Thr Arg 1570 1580
- Lys Val Ser That Asp Asm Ser Leu Thr Leu Arg Gis Asm Giu Asm Giy 535 1590 1595 1600
- Ala Gin Tyr Met Gin Txp G)y Vel Tyr Arg I): Axg Iku Asm Thr Leu 1605 1610 1615
- Fins Ale Ary Glin Leu Val Ale Ary Ale Thr Thr Gly The Asp Thm The 1620 1625 1630
- Leu Ser Met Glu Thr Gln Asn Ile Glu Chu Pro Gln Leu Gly 198 Gly 1835 1640 1645
- Fire Tyr Ala Thr Phe Wal Ile Pro Pro Tyr Aem Pro Ser Thr His Gly 1650 1655 1660
- Amp Glu Arg Trop Fine Lym Leu Tyr Ile Lym Him Val Val Amp Amp Agn 665 1670 1675 1680
- Sar Ris Ile Ile Tyr Ser Gly Gln Leu Lys Asp Thr Asn Ile Ser Thr 1685 1690 1695
- Thr Leu Fhe Ile Fro Leu Asp Asp Val Pro Leu Asm Gln Asp Tyr Ser 1700 1705 1710
- Ala Lys Val Tyr Met Thr Phe Lys Lys Ser Pro Ser Asp Cily Thr Trp 1715 1720 1725
- Trp Cly Pro His Phe Val Ang Asp Asp Tys Cly Ilo Val Thr Yle Asn 1730 1735 1740
- Pro Lys Ser Ile Leu Thr Ris Phe Glu Ser Val Aen Val Leu Aen Aen 745 1750 1755 1760
- The Ser Ser Glu Pro Net Asp Phe Ser GJy Ala Asa Ser Leu Tyr Phe 1765 1770 1775
- Trp Glu Lau Pho Tyr Tyr Thr Pro Met Leu Val Ala Gln Arg Leu Leu 1780 1785 7790
- His Glu Gln Amn Phe Amp Glu Ala Amn Arg Trp Leu Lys Tyr Val Trp 1795 1800 1805
- Ser Pro Sor Cly Tyr Ile Val His Gly Gln 11e Gln Amn Tyr Gln Trp 1810 1815 1820
- Ast. Vol. Ang Pro 180 Leo Clu Asp Thr Ser Trp Asp Ser Asp Pro 180 825 1830 1835 1840
- AMP Ser Vol Amp Pro Amp Ala Val Ala Chr His Amp Pro Met. $9)_{(3)}$ $9y_{(7)}$ 1855
- Tys Val Ser Thr Phe Met Arg Thr Law Asp Law Law IIo Ala Arg Gly 1860 1865 1870
- Asp His Ala Tyr Arm Glo Leo Glo Arg Asp Thr Leo Asm Glo Ala 1875 1885
- Mot Top Tyr Met Gin Ala Leu His Lew Low Gly Asp Dys Pro Tyr Lew

1890	1895	1900	
Pro Leu Ser'	אר יולוני זלוני זלוני זלוני	Asp Pro Arg Leu Asp	lys Ala Ala Asp
905	1.910	1915	1920
He The The C	Gin Ser Ala Rin 8	Ser Ser Ile Val	Ala Leu Arg Gin
	1925	1930	1 93 5
Ser Thr Pro .	Ala Leu Leu Seri	Leu Arg Ser Alm Asn	Thr Low Thr Asp
1	940	1945	1950
Teu Phe Leu		Glu Va l Met Met Aso	'Tyr 'thp Gln Thr
1955		960	965
		Leo Arg His Aso Loo 1980	Ser He Asp Gly

Gin Pro Leu Tyr Leu Pro Ile Tyr Ala Thr Pro Ala Amp Pro Lya Ala 985 1990 1995 2000

Glu Ser Who Mot Ser Lew Trp Arm The Pro His Mot Lew Glu Amn Ala 2020 2025 2030

Army Ser Met Val Ser Gln Leu Thr Gln Phe Gly Ser Thr Leu Gln Asn 2035 2040 2045

lle lle Glu Arg Gl
n Asp Ala Glu Ala Leu Asp Ala Leu Glu Asp 2050 2060

Gln Ala Ala Glu Lau Ile Lau Thr Asn Leu Ser Ile Gln Asp Lys Thr 065 2070 2075 2080

The Glu Glu Leu Asp Ala Glu Lys Otr Val Leu Glu Lys Ser Lys Ala 2065 2090 2095

Gly Ala Gln Ser Arg Phe Asp Ser Tyr Ser Lyo Ice His Asp Glo Asn 2100 2105

lle Asn Ala Cly Glu Asn Gln Ala Met Thr Leu Arg Ala Ser Ala Ala 2115 2120 2125

Gly Leu Thr Thr Ala Val Glm Ala Ser Arg Leu Ala Gly Ala Ala Ala 2130 2140

Asp Leu Val. Pro Ast Illo Pho Cly Pho Ala Cly Cly Cly Ser Arg Trp 145 2150 2155 2160

Gly Ala Ile Ala Glu Ala Thr Gly Tyr Val Wet Glu Pho 9xr Ala Asn 2175 2175

Val Not Asn Thr Glu Ala Asp Lyo Ile Ser Glu Ser Glu Thr Tyr Arg 2180 2185 2190

Arg Arg Glu Glu Trp Glu Ite Glu Arg Asu Asu Ala Glu Ala Glu 2195 2200 2205

leu bys Gin Lou Asp Ala Gin Leu Lys Ser Leu Ala Val Ang Ang Giu 2210 2215 2220

Ala Ala Val Leu Glo Lys Thr Ser Leu Lys Thr Glo Glo Glu Glo Thr 225 2230 2235 2240 Gin Ala Gin Lou Ala Pho Lou Gin Arg Lys The Ser Ash Gin Ala Leu 2265 2250 2255

Tyr Asn Trp Len Arg Gly Arg Len Ale Ale Ilo Tyr Hos Gln Phe Tyr 2260 2265 2270

Asp Leu Ala Ilo Ala Arg Oys Leu Met Ala Glu Gln Ala Tyr Arg Trp 2275 2280 2285

Glu lle Ser Asp Asp Ser Ala Arg Fhe Ile Lys Pro Gly Ala Trp Gln 2290 2295 2300

Gly Thr Tyr Ala Gly Leu Leu Ala Gly Glu Thr Leu Mot Leu Ser Leu 305 2310 2315 2320

Ala Gin Met Glu Asp Ala His Lou Arg Asp Lys Arg Ala Leo Glu 2325 2330 2335

Val Glu Arg Thr Val Ser Leo Ala Glu Ile Tyr Ala Gly Leo Pro Gln 2340 2345 2350

Amp Lym Gly Pro Pine Ser Leu Thr Glπ Clu Ilo Glu Lym Leu Val Am 2355 2360 2365

Ale Gly Ser Gly Ser Ale Gly Ser Gly Asn Asn Leo Ale Pho Gly 2370 2380

Ala Gly Thr Asp Thr Lys Thr Ser Leu Cln Ala Ser Ile Ser Leu Ala 385 2390 2395 2400

Arg Ile Lyn Gin Ile Ser Val Thr Leu Pro Ala Leu Leu Giy Pro Tyr 2420 2430

Gin Asp Vai Gin Ala The Leu Ser Tyr Gly Asp Lys Ala Gly Leu Ala 2435 2440 2445

Asin Gly Cyb Ala Ala Lou Ala Vol Ser His Gly Thr Asin Aqu Sor Gly 2450 2455 2460

Cln Fire Gln Leu Asp the Asn Asp Gly Lys Pho Leu Pro the Glu Gly 465 2470 2475 2480

lle Ala Ile Asp Glu Gly Thr Leu Thr Leu Ser Phe Pro Ash Ala Ser 2485 2490 2495

The Pro Ala Lye Gly Lys Glm Ala The Met Lew Lye The Low Asn Aep 2500 2505 2510

lle lle leu His Ile Ang Tyr Thr Ile Lys $2515 \hspace{1.5cm} 2520 \hspace{1.5cm}$

<210> 14

<211> 1481

<212> PRT

:213: Photortobdus luminescers:

<400> 14

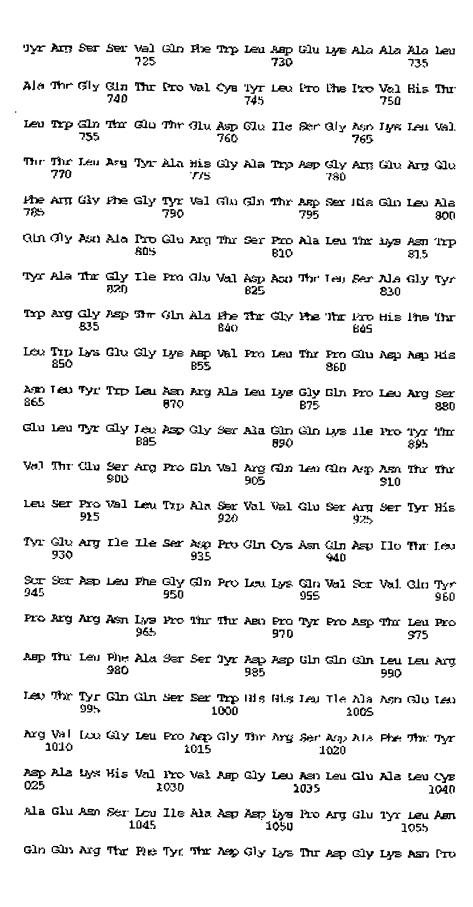
Met Gin Asn Ser Gin Thr Pho Ser Met Thr Glu Leu Ser Leu Pro Lys 1 5 10

Gly Gly Gly Ala 11e Thr Gly Met Gly Glu Ala Leu Thr Pro Ala Gly



			20					25					30		
Pro	Авр	G ly 35	Met,	a la	Ala	Leu	Ser 40	Leu	Pro	Leu	Pro	Ile 45	Ser	Ala	Gly
Arg	Gly 50	Ίγτ	Ala	Pro	Ser	Leu 55	Trac	Lou	Asn	Тух	Asn 60	Ser	Gly	Thr	Gly
Asn 65	Ser	Fro	Fine	Gly	Барід 70	Gly	Тър	ABP	CAE	Asn 75	Val	Meet.	TÌ	T.Le	Л ту 80
Ang	Arg	Thr	Ser	Thr 85	Gly	Val	Pro	Aeto	ፓ ሃድ 90	ሊ ሚን	Glu	An	Assp	Thr 95	Pho
Leu	Gly	PTO	Glu 100	GЈУ	G).u	Val	Leu	Val 105	Val	Ala	Leu	Jesin	Ghi 110	Ala	яy
្សា	Ala	Дэр 115	ıle	Arg	Sen:	ផាររ	8ar 120	Seor	Leu	Gln	Gly	Ile 125	Aen	Leu	Gly
Met	Tîm 130	Phe	Tîur	Val	The	Gly 135	тут	Arg	Ser	Arg	Lен 140	GI u	Sed	His	Stoc
Ser 145	Arg	Len	Glu	Тут	117p 150	Gln	Pro	Gln.	Thr	Tha 158	Gly	Aka	'lfur	ASP	9he 160
Тхр	Lau	Ile	Tyr	Ser 165	Pχυ	Asp	Gly	CJD	Ala 170	}tis	Lev	Ley	മുഴ	Lyrs 17 5	Asn
Pro	Gľu	λ]а	Arg 180	Tle	Ser	Aen	Pro	Leu 185	Aso	Val	Aeto	CJD	Thr 190	Ala	വ്വാ
Тъ	Leu	Leu 195	Glu	Mle	Sex	Val	Ser 200	Ser	His	Gly	Olu	01n 205	Ha	Tyr	Тут
Gln	Тут 216	Ary	Ala	Glu	ABP	Glu 215	The	డిణ్లు	Cya	Glu	Thr 2 2 0	λep	Glu	Leu	Thr
Ala 225	Hís	Pro	лап	Thur	9 ta r 230	Val	Gln	Ing	Тухг	Leu 235	Gln	Val	Val	Ris	1 y r 240
αГА	Asn	Leu	Thar	Ala 2 4 5	Ser	Glu	Val	Phe	Pro 250	Thr	Leu	Asn	Зlу	Дар 255	Авр
Pro	Leu	Lys	Ser 260	αly	Top	Leu	Phe	∩y≊ 265	Len	Val	P?re-	Анцэ	ፕ ሃድ 270	Gly	Glu
Arg	TÀS	Лет і 275	Ser	Leu	Ser	Glu	M et 280	Pro	Pro	Phe	Lув	Ala 285	TÎTE	Ser	Agn
Jxp	Leu 290	റുജ	Αrg	Lya	Asp	л rg 295	Pha	Ser	Arg	Tyr	Glu 300	<u></u> ያንን	Gly	Phe	Ale
Leu 305	yng	ጥェ	Arg	Arg) թա 310	Cys	Arq	GГи	Ile	1491 335	Met,	Pho	Ris	Arg	Leu 320
Gln	Ul DE	راضا	Ser	Gly 325	GŢu	Ala	lys	Gly	Авр 330	Авр	Glu	Pho	Ala	1 <i>e</i> u 335	Val
8र.च	Arg	Leu	11⊕ 340	Leu	Asp	Тут	Asp	Glu 345	Asn	Ala	Val	Val	Ser 350	TÎU	Leu
Val	Ser	val 355	Ахц	Arg	Val	Gly	н <u>ј.</u> s 360	Clu	CJU	Азр	Сĵу	19u 365	Tîtu	Аľв	Val

Ala	Lou 370	Pro	Pro	Leu	ផ្សារ	Leu 375	Ala	ľуг	ĠĴΩ	Pro	Phe 380	Glu	Pro	Glu	Œυ
Lys 385	Ala	Leni	Tip	Arg	PTO 390	Xet	лар	Va:1	Leu	Ala 395	Aso	the	Asn	Thr	Ile 400
Glo	Arg	qrľ	Gln	Leu 405	Letta	ÀSP	Leu	Gln	Gly 420	Glu	З	Vəl	Prta	Gly 415	lle
Leu	ΊΥΤ	Gln	Авр 420	Lys	Asm	Glγ	Trp	Trp 425	Tyr	Àτυ	Sec	Ala	Gln 430	Arts	ദ്വാ
Thr	Gly	Glu 435	Glu	Met	Aso	Ale	Val 440	Thur	Tip	Gly	TÀS	Me t 44 5	Gìn	Leu	Le≱t)
Pro	Ile 450	Thr	Pro	Ala	Ίle	Gln 455	ASP	Asn	Ala	Stor	Leu 460	Met	ABP	Ile	Asm
61y 465	Asp	Gly	Gln	Let)	1ед) 470	Trp	Val	Ile	Thr	வழ 475	Pro	Gly	Leu	Airg	Gly 480
	Hie			485					490					495	
	Ala		500					505					510		_
Leo	Met	Gly 515	Ala	Gly	leu	Ser	Asp 520	Leu	Val	Leu	Ile	Gly 52 5	Pno	Lys	Secr
Val	Arg 530	Leu	Тут	Ala	Le n	Asn 535	Arg	Λεp	αlγ	Ite	1hr 540	Glu	Gly	Arg	A6 p
Val 545	Val	Glŋ	Ser	αјλ	Gly 550	Ile	Thu	Leu	PTO	Leu 555	Pro	αly	Ale	Asp	Ala 560
Arg	Lys	Lex	Val	Ala 565	Pino	Secr	Aep	Val	1 <i>e</i> u 570	Оίγ	Ser	Gly	Gln	Ala 575	His
Leu	Val	Glu	Val 580	Ser	Ala	Thr	Lys	Val 565	Thar	Cys	тър	ያካን	Asa 590	Lou	GLy
His	Gly	Arg 595	Phe	GLy	Gln	Pro	T)e 600	The	Leo	Pro	αlу	9he 605	Ser	Gln	Ser
Ala	ለጋል 610	Aan	Phe	Asar	Pro	Аяр 615	Arg	Val.	His		Ala 620		Leu	ХВ Ъ	gly
Ser 625	Gìy	Pro	Ala	Asp	Le n 630	Ţle	Тут	۷al	Hds	Мв 635	እሟን	Юs	Lox	Aap	Tle 640
Phe	Ser	Aso	Gl 13	Ser 645	G1y	Aan	Сìу	Phe	Ala 650	СIл	Pro	Phe	Tra	Leu 655	Arg
Phe	Pro	PS	Gly 660	kou	Àrg	Phe	AB ₂)	љ _{ж)} 665	Ttn-	Сув	Gln	نها	Gln 670	Val	Ala
УВД	Va l	Cln 675	Gly	Legy	Cly	Val	Val 680	Sea:	T e ja ju	IJ.c	Lew	Ser 685	Val	Pro	Нis
Met	Ala 690	Pro	His	Hie	Ттр	Arg 695	Cys	Аер	Leu	Thr	Ле п 700	Ala	Lys	PTO	Тър
1.eu 705	Leu	Ser	ദിവ	Met	Asn 710	Asn	Aan	Met.	Gly	л]д 715	His	His	Thur	Leu	His 720



20	ONI	1065		1070
			-22 -2	
1075	-A2 11W 1-1C	1080	Ala Leu Ile Ala 108:	
Thr Ala Val I 1090	L eu Thr Glu	Ser Leu leu : 1095	Ser Alo Ita Asp 1100	o Gly Gly Ile
Thi: Pro Asp (105	Flu Leu Pro 1110	Gly Len Len (Tto: Gln Ala Gly 1115	/ Tyr Gln Gln 1120
Glu Inn Tyr I	Jeu Phe Pro 1325	- Lea Set (Rly (1:	Glu Ası Glu Val 130	l Tip Val Ala 1135
Arg Lys Gly 1	lyn Thr Asp 140	Tyr Gly Thr (1145	Glu Val Gln Phe	Trp Arg Pro 1150
Val Ala Gln A 1155	arg Asn The	Gln Leu Thr (1.160	Gly Lys Tho Tho 1369	
App The His 1 1170		ile എന്ന ദിന എ 1175	Thr Gin Ag, Ai/ 1180	a Ale Gly Lev
Thr Val Ser 8 185	Ча Ав п Тул 1190	Asp Trp Arg :	Phe Leu Thr Pro 1195	> Met Gln Leo 1200
Thr Asp Ile ?	Aso Asp Aso 1205	.Val dis Ile : 13	Ile Thr Len Asp 210	Ale Leo Gly 1215
Arg Pro Val 1	ffor Gla Arg 220	The Trp Cly : 1225	Ile Glu Amn Gly	y Val Ala Thr 1230
Gly Tyr Ser 8 1235	er iro Glu	Ala Lye Pro (1240	Phe Thr Pro Pig 1209	
Asin Ala Ala 1 1250	lle Ala Leo	Thr Cly Pro 1 1255	Lew Pro Val Ala 1260	a Gln tys Lev
Val Tyr Ala F 265	Pro Amp Ser 1270	Trp Met Pro 1	Leu Phe Gly Gir 1275	Glu Thr Phe 1280
Aso The Leu 1	Ռու Յիր Յիս 1285		Thir Leu Arg Asq 290	Len Arg Ile 1295
Ilo Thr Glu A	Map Trop Arg 300	Ile Cys Ala 1 1305	Leu Ala Arg Arç	j Arg Trp Leu 1310
Gln Ser Gin I 1315	we Ala Gly	Thr Pro Teu 1 1320	Val Lys Lou Lo. 1329	
ile Gly Len p 1330		Aen Leu Xet 1 1335	Leo Ala Thr Asg 1340	Ang Tyr Asp
Ang Amp Ser 6 345	ປັນ Gl n Gln 1350	Ile Arg Glm (Gln Val Ala ehe 1355	e Ser Asp Gly 1360
Phe Gly Arg 1	æu Leo Gin 1365		Arg Bia Glo Ala 370	Gly Glu Ala 1375
Trp Gln Ary A	car Glu Vab	Gly Sen Len 1 1385	Val Thr lys Mot	: Gl.u Asp Thr 1390
Lys Thr Arg T 1395	m ala Ilc	Thr Gly Arg 9	Thr Glu Tyr Asp 1409	

Gln Ala 11e Arg Tir Tyr Gln Pro Tyr Phe Leu Arn Arp Trp Arg Tyr 1410 1415 1420	
Vel Ser Amp Amp Ser Ala Arg Lye Glu Ala Tyr Ala Amp Thr Him The 425 1430 1435 1440	
Tyr Asp Pro Ile Gly Arg Glo Ilo Cla Val Ilo Thr Ala Lys Gly Trp 1445 1450 1455	
Leu Arg Gln Aen Gln Tyr Phe Pro Trp Phe Thr Val Ser Glu Aep Glu 1460 1465 1470	
Asn Asp Leu Ser Ala Asp Ala Leu Val 1475 1,470	
<210> 15 <211> 23 <212> DNA <213- Artificial Sequenco	
<220> <223> Description of Artificial Sequence: Oligonucleotide	
<600> 15 egggateega tgettiitaaa agg	23
<210> 16 <211> 16 <212> DNA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonucleotide	
<400> 16 gogpcrattgs tittgag	16
<pre><210> 17 <211> 19 <212> INA <213> Artificial Sequence</pre>	
<220> <223> Description of Artificial Sequence: clipsonuclootide	
<400> 17 cattamagn cyanognac	19
<210> 18 <213> 26 <212> INA <213> Artificial Sequence	
<220> <223> Description of Artificial Sequence:oligonocloptiche	
<400> 18 gagagagate ttachtaatg gtg(ag	26
<210> 19	

<211> 28	
<212> INB	
<213> Artificial Sequence	
<220>	
<223> Description of Artificial Seguence:oligonucleotide	
<400> 19	
ongogageta catgongaat tracagae	28
<210> 20	
<211> 18	
<212> INA	
<pre><213> Artificial Sequence</pre>	
<220>	
<223> Description of Artificial Sequence:oligorucleotide	
<400> 20	
ggceathgca g cpataa g	19
<210> 21	
<211> 18	
<212> DNA	
<213> Artificial Sequence	
William Sayette	
<220>	
<223> Description of Artificial Sequence:oligonuclectide	
<400> 21	
cetraacgos ggaagago	18
<210> 22	
<211> 26	
<211> 20 <212> DNA	
<213> Artificial Sequence	
<220>	
<223> Description of Artificial Sequence:oligonucleotide	
<400> 22	
geoctogagt tacacqagog cytoag	26

		,
		٠

PCT





INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6:

C12N 15/31, 15/82, 15/10, 1/21, 5/10, A01H 5/00, C07K 14/24, A01N 63/02 **A3**

(11) International Publication Number:

WO 99/42589

(43) International Publication Date:

26 August 1999 (26.08.99)

(21) International Application Number:

PCT/EP99/01015

(22) International Filing Date:

18 February 1999 (18.02.99)

(30) Priority Data:

09/027,080 60/116,439 20 February 1998 (20.02.98) US

20 January 1999 (20.01.99) US

(71) Applicant (for all designated States except AT US): NOVAR-TIS AG [CH/CH]; Schwarzwaldallee 215, D-4058 Basel (CH).

(71) Applicant (for AT only): NOVARTIS-ERFINDUNGEN VER-WALTUNGSGESELLSCHAFT MBH [AT/AT]; Brunner Strasse 59, A-1235 Vienna (AT).

(72) Inventors; and

(75) Inventors/Applicants (for US only): KRAMER, Vance, Cary [US/US]; 608 Dana Court, Hillsborough, NC 27278 (US). MORGAN, Michael, Kent [US/US]; 5805 Garrett Road, Durham, NC 27707 (US). ANDERSON, Ame, Robert [US/US]; 1005 Green-Pace Road, Zebulon, NC 27597 (US). HART, Hope, Prim [US/US]; 4106 Planters Glen Court, Fuquay-Varina, NC 26526 (US). Warren, Gregory, Wayne [US/US]; 324 Bond Lake Drive, Cary, NC 27513 (US). DUNN, Martha, M. [US/US]; 6201 Oakview Court,

Hillsborough, NC 27278 (US). CHEN, Jeng, Shong [-/US]; 302 Orchard Lane, Chapel Hill, NC 27514 (US).

(74) Agent: BECKER, Konrad; Novartis AG, Corporate Intellectual Property, Patent & Trademark Dept., CH-4002 Basel (CH).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, Cl, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report:

23 December 1999 (23.12.99)

(54) Title: INSECTICIDAL TOXINS FROM PHOTORHABDUS

(57) Abstract

Novel nucleic acid sequences isolated from Photorhabdus luminescens, whose expression results in novel insecticidal toxins, are disclosed herein. The invention also discloses compositions and formulations containing the insecticidal toxins that are capable of controlling insect pests. The invention is further drawn to methods of making the toxins and to methods of using the nucleotide sequences, for example in microorganisms to control insect pests or in transgenic plants to confer insect resistance.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL.	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	ТJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
ВJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of Americ
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JР	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

INTERNATIONAL SEARCH REPORT

Inter- itional Application No 01015 Pui/EP

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/31 C12N15/82

A01H5/00 C07K14/24 C12N15/10 A01N63/02

C12N1/21

C12N5/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols) I PC 6 C12N A01H C07K A01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUM	NTS CONSIDERED TO BE RELEVANT	
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 97 17432 A (WISCONSIN ALUMNI RES FOUND) 15 May 1997 (1997-05-15) the whole document, particularly SEQ ID NOS 31,46,47,48,49,50,51,60	1-3,7-9, 11-24, 26-36
P,X	WO 98 08932 A (DOW AGROSCIENCES LLC; WISCONSIN ALUMNI RES FOUND (US)) 5 March 1998 (1998-03-05) see pages 209-210,215-224,231-237, and 243-245.	1-3,7-9, 11-24, 26-36

X Further documents are listed in the continuation of box C.	Patent family members are listed in annex.		
Special categories of cited documents : A* document defining the general state of the art which is not considered to be of particular relevance.	T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention		
"E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "&" document member of the same patent family		
Date of the actual completion of the international search	Date of mailing of the international search report		
20 October 1999	0 8. 11. 99		
Name and mailing address of the ISA	Authorized officer		
European Patent Office, P.B. 5818 Patentiaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo ni, Fax: (+31-70) 340-3016	Maddox, A		

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Int	tional Application No	
	T/EP 99/01015	

	CITE 33/01013						
C.(Continua Category °	ction) DOCUMENTS CONSIDERED TO BE RELEVANT Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.					
A	DAVID JOSEPH BOWEN: "Characterization of a High Molecular Weight Insecticidal Protein Complex Produced by the Entomopathogenic Bacterium Photorhabdus luminescens (Nematodes, Biological Control)" THESIS UNIVERSITY WISCONSIN, 1 May 1995 (1995-05-01), XP002076022 see chapter 3	1-36					
A	WO 95 00647 A (COMMW SCIENT IND RES ORG;SMIGIELSKI ADAM JOSEPH (AU); AKHURST RAY) 5 January 1995 (1995-01-05) the whole document	1-36					
Α	SZITTNER, R., ET AL.: "Nucleotide sequence, expression, and properties of luciferase coded by the lux genes from a terrestrial bacterium" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 265, no. 27, 1990, pages 16581-16587, XP002119674 figure 5	2,11					
A	WO 93 07278 A (CIBA GEIGY AG) 15 April 1993 (1993-04-15) the whole document	12-19, 29-34					
P,A	WO 98 08388 A (MORGAN JAMES ALUN WYNNE ;JARRETT PAUL (GB); ELLIS DEBORAH JUNE (GB) 5 March 1998 (1998-03-05) see SEQ ID NO:1	1-36					



PCT/EP 99/ 01015

Box I Obs rvations wher certain claims were found unsearchable (Continuation fitem 1 of first shiet)
This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:
Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
2. Claims Nos.: because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. X Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
B x II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)
This International Searching Authority found multiple inventions in this international application, as follows:
see additional sheet
As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Remark on Protest The additional search fees were accompanied by the applicant's protest. X No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 4,5,6,10,25 all completely, and 1-3,12-24, 27-36 all partially

Nucleic acid molecule comprising the claimed regions of sequence ID 1, chimeric genes and hosts containing said molecule, toxins expressed by said regions, and method for producing said toxins and controlling insects using said toxins, method for mutagenizing said nucleic acid molecules.

2. Claims: 7-9,11,26 all completely, and 1-3,12-24, 27-36 all partially

Nucleic acid molecule comprising the claimed regions of sequence ID 11, chimeric genes and hosts containing said molecule, toxins expressed by said regions, and method for producing said toxins and controlling insects using said toxins, method for mutagenizing said nucleic acid molecules.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210
Continuation of Box 3.
The reference to claim 44 in claim 30 is inconsistent with the numbering of the claims, since claim 44 has not been filed. For the purpose of defining the search, claim 30 has been considered to refer to the toxin of claim 20, and searched accordingly.

INTERNATIONAL SEARCH REPORT

Intr tional Application No

	tent document i in search report		Publication date		atent family nember(s) _	Publication date
WΩ	9717432	A	15-05-1997	AU	1050997 A	29-05-1997
	J. 21 14E		·	BR	9606889 A	28-10-1997
				CA	2209659 A	15-05-1997
				EP	0797659 A	01-10-1997
				HU 	9900768 A	28-06-1999
				PL	321212 A	24-11-1997
				SK	93197 A	06-05-1998
•				AU	2829997 A	19-03-1998
				WO	9808932 A	05-03-1998
wn	9808932	 A	05-03-1998	AU	1050997 A	29-05-1997
	,000,02	.,		AU	2829997 A	19-03-1998
				BR	9606889 A	28-10-1997
				CA	2209659 A	15-05-1997
				EP	0797659 A	01-10-1997
				HU	9900768 A	28-06-1999
				PL	321212 A	24-11-1997
				ŠK	93197 A	06-05-1998
				WO	9717432 A	15-05-1997
wo.	9500647		05-01-1995	AU	675335 B	30-01-1997
","		• •		AU	6991694 A	17-01-1995
				EP	0705340 A	10-04-1996
				JP	9500264 T	14-01-1997
WO	9307278	Α	15-04-1993	US	5625136 A	29-04-1997
				ΑU	2795292 A	03-05-1993
				BG	98747 A	28-02-1995
				BR	9206578 A	11-04-1995
				CA	2120514 A	15-04-1993
				CZ	9400769 A	15-03-1995
				EP	0618976 A	12-10-1994
				HU	68261 A	28-06-1995
				JP	7500012 T	05-01-1995
				RO	110263 A	30-11-1995
				SK	37894 A	05-10-1994
				US	5859336 A	12-01-1997
WO	9808388	Α	05-03-1998	AU	4024997 A	19-03-1998
				EP	0923295 A	23-06-1999